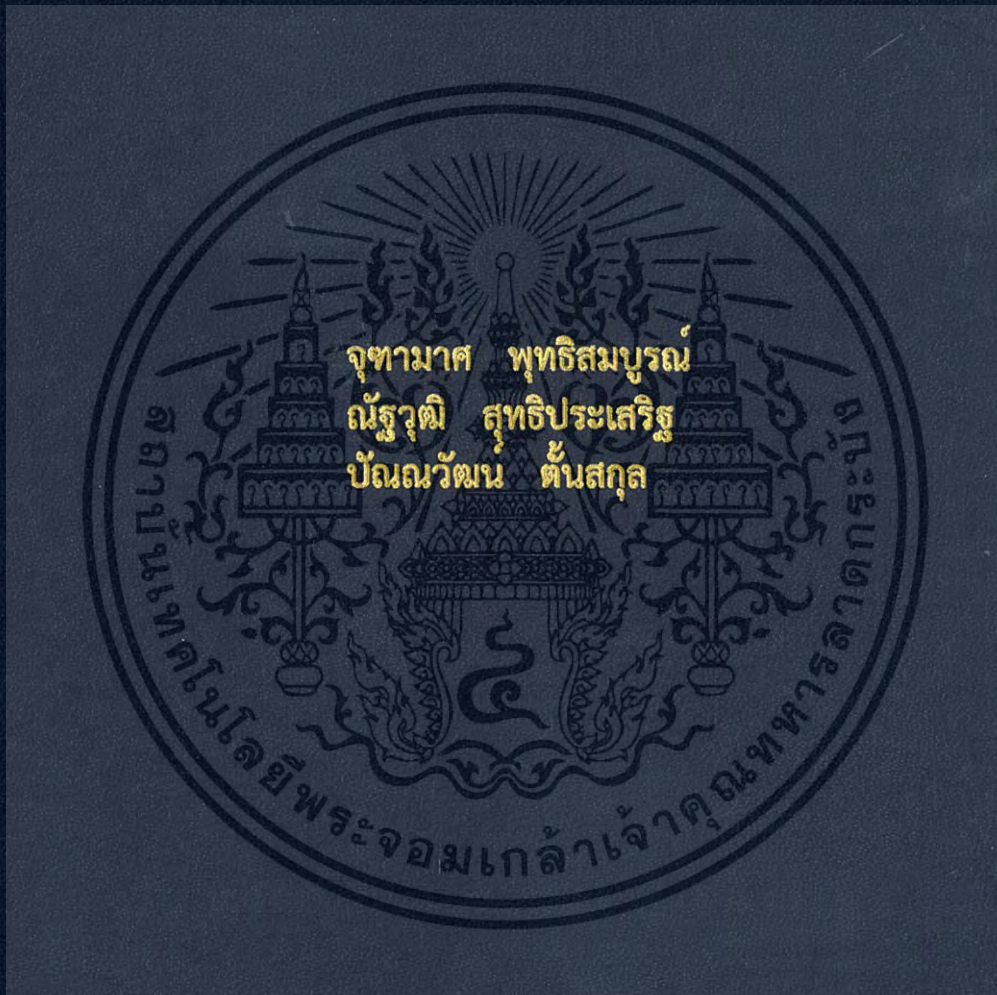


ระบบอัจฉริยะในการค้นหาโรงแรม

INTELLIGENT HOTEL SEARCHING SYSTEM



ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2558

ระบบอัจฉริยะในการค้นหาโรงแรม

INTELLIGENT HOTEL SEARCHING SYSTEM



T149253



จ 628 ร

สงทญ... 2558
เลขทะเบียน... 149253
วันเดือนปี... 30 อ.ค. 2561

b. 149253/16
i.

ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2558

INTELLIGENT HOTEL SEARCHING SYSTEM



A SPECIAL PROBLEM SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENT FOR
THE DEGREE OF BACHELOR OF SCIENCE (COMPUTER SCIENCE)
DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
ACADEMIC YEAR 2015

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ ระบบอัจฉริยะในการค้นหาโรงแรม
Intelligent Hotel Searching System

ชื่อนักศึกษา นางสาวจุฑามาศ พุทธิสมบุรณ์ รหัสนักศึกษา 55050252
นายณัฐวุฒิ สุทธิประเสริฐ รหัสนักศึกษา 55050298
นายปณณวัฒน์ ต้นสกุล รหัสนักศึกษา 55050371

ปริญญา วิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชา วิทยาการคอมพิวเตอร์
ปีการศึกษา 2558
อาจารย์ที่ปรึกษา ผศ.ดร.อนันตพร หารัชคณาชัย

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง(สจล.) อนุมัติให้
ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต (วิทยาการ
คอมพิวเตอร์) ประจำปีการศึกษา 2558

คณะกรรมการสอบ	ลายมือชื่อ
ดร.สายชล ใจเย็น ประธานกรรมการ	
ดร.สันติภรณ์ นรบิน กรรมการ	
ผศ.ดร.อนันตพร หารัชคณาชัย กรรมการและอาจารย์ที่ปรึกษา	

ลิขสิทธิ์ของคณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ	ระบบอัจฉริยะในการค้นหาโรงแรม		
ชื่อนักศึกษา	นางสาวจุฑามาศ พุทธิสมบุรณ์	รหัสนักศึกษา	55050252
	นายณัฐวุฒิ สุทธิประเสริฐ	รหัสนักศึกษา	55050298
	นายปณณวัฒน์ ตันสกุล	รหัสนักศึกษา	55050371
ปริญญา	วิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)		
ภาควิชา	วิทยาการคอมพิวเตอร์		
คณะ	วิทยาศาสตร์		
มหาวิทยาลัย	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง(สจล.)		
ปีการศึกษา	2558		
อาจารย์ที่ปรึกษา	ผศ.ดร.อนันตพร หารราชคุณาฒย์		

บทคัดย่อ

ปัจจุบันอุตสาหกรรมการท่องเที่ยวในประเทศไทยได้เจริญเติบโตส่งผลให้จำนวนโรงแรมที่พักเพิ่มขึ้น เว็บไซต์จึงเป็นอีกช่องทางหนึ่งที่ใช้ในการประชาสัมพันธ์ข้อมูลเกี่ยวกับโรงแรม ซึ่งมีบางเว็บไซต์ที่รวบรวมข้อมูลห้องพัก บางเว็บไซต์เปิดให้ผู้ใช้มีส่วนร่วมในการแสดงความคิดเห็น ระบบนี้ได้ประยุกต์ใช้เทคนิคการจำแนกประเภทแบบเบย์อย่างง่ายเพื่อช่วยในการค้นหาโรงแรม โดยข้อมูลฝึกสอนที่นำมาใช้ประกอบด้วยข้อมูลความคิดเห็น 1800 ความคิดเห็นของโรงแรม 60 แห่งในเขตการปกครองพิเศษเมืองพัทยา จังหวัดชลบุรี ผลลัพธ์ที่ได้จากโมเดลการค้นหาจะแสดงผลบนหน้าเว็บแอปพลิเคชันตามลำดับรายชื่อโรงแรมจากคำค้นหาของค่าความน่าจะเป็นที่ได้จากตัวจำแนกแบบเบย์อย่างง่าย โดยผู้ใช้สามารถค้นหาข้อมูลโรงแรมได้ตรงตามความต้องการมากที่สุด

คำสำคัญ: การค้นหาโรงแรม การทำเหมืองข้อมูลความคิดเห็น ตัวจำแนกแบบเบย์อย่างง่าย

Title	Intelligent Hotel Searching System			
Students	Miss Jutamas	Puttisomboon	Student ID	55050252
	Mr. Natthawud	Sutthiprasert	Student ID	55050298
	Mr. Pannavat	Tansakul	Student ID	55050371
Degree	Bachelor of Science (Computer Science)			
Department	Computer Science			
Faculty	Science			
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)			
Academic Year	2015			
Advisor	Asst.Prof.Dr. Anantaporn Hanskunatai			

Abstract

Nowadays, travel industrial in Thailand has developed and consequently affects the rising up of hotels and residences. Website is one way to access the information of hotels. There are some websites that conclude all information and user can give opinions. This system is applied Naïve Bayes classifier to assist in searching for any hotels. The datasets were collected from 1,800 reviews of 60 hotels in Pattaya, Chonburi province. The result of the model will appear in a list of hotel names ranking on the web application from probability related to user keywords by Naïve Bayes classifiers which any users can search appropriate hotels according to their requirements.

Keywords : Hotels searching, Opinion mining, Naïve Bayes classifier

กิตติกรรมประกาศ

หัวข้อปัญหาพิเศษนี้สำเร็จลุล่วงลงได้ด้วยความกรุณาจากบุคคลผู้มีพระคุณหลายท่านคณะผู้จัดทำจะขอกล่าวขอบคุณดังรายละเอียดด้านล่าง

ขอขอบคุณ ผศ.ดร.อนันตพร ทรราชคุณาตย์ อาจารย์ที่ปรึกษาที่ให้ความรู้ แนะนำแนวทางในการแก้ปัญหาและข้อบกพร่องที่เกิดขึ้น รวมทั้งให้ความช่วยเหลือ เสียสละเวลาเพื่อให้คำปรึกษาอย่างสม่ำเสมอ

ขอขอบคุณ ดร.สายชล ใจเย็นและดร.สันติภูษณ์ นรบิน ที่สละเวลามาเพื่อให้คำแนะนำ แนวทางการแก้ปัญหาต่างๆที่ทำให้งานสมบูรณ์ขึ้น

ขอขอบคุณอาจารย์ทุกท่านในภาควิชาวิทยาการคอมพิวเตอร์ที่อบรมสั่งสอนให้ความรู้ให้แก่คณะผู้จัดทำสำหรับการศึกษาในระดับปริญญาตรี จนกระทั่งปัญหาพิเศษนี้ได้สำเร็จลุล่วงไปด้วยดี

สุดท้ายขอขอบคุณบิดา มารดาเป็นอย่างสูงที่ได้ให้ความกรุณาและเป็นกำลังใจช่วยเป็นแรงผลักดันตลอดการศึกษา



จุฬามาศ

ญัฐวุฒิ

ปณณวัฒน์

พุทธิสมบูรณ์

สุทธิประเสริฐ

ต้นสกุล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ.....	ง
สารบัญตาราง.....	ฉ
สารบัญรูป.....	ช
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญ.....	1
1.2 วัตถุประสงค์ของปัญหาพิเศษ.....	1
1.3 ข้อจำกัดและขอบเขตของปัญหาพิเศษ.....	1
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	2
1.5 ขั้นตอนในการดำเนินการ.....	2
1.6 อุปกรณ์ที่ใช้ในการทำปัญหาพิเศษ.....	2
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	3
2.1 การทำเหมืองข้อมูล.....	3
2.1.1 ความหมายของการทำเหมืองข้อมูล.....	3
2.1.2 ขั้นตอนการทำเหมืองข้อมูล.....	3
2.1.3 เทคนิคการทำเหมืองข้อมูล.....	5
2.1.4 ประโยชน์ของการใช้เหมืองข้อมูล.....	5
2.1.5 แนวโน้มและการประยุกต์ใช้งานเหมืองข้อมูล.....	6
2.2 การจำแนกประเภทข้อมูล.....	6
2.3 การจำแนกประเภทแบบเบย์อย่างง่าย.....	7
2.4 เทคนิคการคัดเลือกคุณลักษณะ.....	9
2.5 มาตรวัดความสอดคล้องของเคนดอลล์.....	10
2.6 งานวิจัยที่เกี่ยวข้อง.....	12
2.6.1 การวิเคราะห์ความรู้สึกของสถานะบนเฟสบู๊คด้วยตัวจำแนกประเภทแบบเบย์ ..	12
2.6.2 การจำแนกแบบเบย์เป็นการจำแนกประเภทเอกสารที่ดีที่สุด.....	14

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
บทที่ 3 โมเดลค้นหาโรงแรมและการออกแบบระบบ.....	16
3.1 การสร้างโมเดลค้นหาโรงแรม	16
3.1.1 ขั้นตอนการเตรียมข้อมูล	16
3.1.2 ขั้นตอนการสร้างโมเดล	22
3.1.3 ขั้นตอนการใช้งานโมเดล	23
3.2 การวิเคราะห์และออกแบบระบบ	24
3.2.1 แผนภาพการทำงานของผู้ใช้งาน (Use Case Diagram)	24
3.2.2 ผังงาน (Flow Chart Diagram).....	26
3.2.3 แผนภาพแสดงลำดับการทำงานของระบบ (Sequence Diagram).....	28
3.2.4 การออกแบบฐานข้อมูล.....	29
3.2.5 ตัวอย่างของส่วนติดต่อกับผู้ใช้ (Graphic User Interface).....	31
บทที่ 4 การทดลอง ผลการทดลองและการพัฒนาเว็บแอปพลิเคชัน.....	34
4.1 การทดลองและผลการทดลอง.....	34
4.1.1 การทดลอง.....	34
4.1.2 ผลการทดลอง	35
4.2 การพัฒนาเว็บแอปพลิเคชัน	36
4.2.1 ส่วนของหน้าจอแสดงผล	36
บทที่ 5 บทสรุปและข้อเสนอแนะ.....	42
5.1 บทสรุป	42
5.2 ข้อเสนอแนะ	42
เอกสารอ้างอิง	43
ภาคผนวก.....	44
ภาคผนวก ก.....	45
ภาคผนวก ข.....	49

สารบัญตาราง

ตารางที่	หน้า
2.1 อันดับที่แท้จริงเทียบกับอันดับที่เกิดจากการทำนาย.....	11
2.2 ค่าความสอดคล้องกันของแต่ละคู่ลำดับของผลลัพธ์.....	11
3.1 ตัวอย่างการสร้างรากศัพท์ของคำ	20
3.2 ตัวอย่างการแปลงข้อมูลความคิดเห็น.....	21
3.3 ค่าความน่าจะเป็นในการปรากฏคำศัพท์ของแต่ละโรงแรม.....	22
3.4 ตัวอย่างความน่าจะเป็นด้านความสะอาด.....	23
3.5 ตัวอย่างผลลัพธ์ที่ได้จากการเรียงลำดับ	24
3.6 รายละเอียดของตาราง Hotels.....	30
3.7 รายละเอียดของตาราง Rate	31
3.8 รายละเอียดของตาราง Hotels_pic.....	31
4.1 ชุดคำค้นหาเพื่อใช้ในการทดสอบประสิทธิภาพ	35
4.2 ค่าเคนดอลล์ที่ได้จากโมเดลการคัดเลือกคุณลักษณะด้วยเทคนิค ค่าสารสนเทศเกิน 5% ถึง 50%.....	35
4.3 ค่าเคนดอลล์ที่ได้จากโมเดลการคัดเลือกคุณลักษณะด้วยเทคนิค ค่าสารสนเทศเกิน 55% ถึง 100%.....	36

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 ขั้นตอนการทำเหมืองข้อมูล	5
2.2 ขั้นตอนการสร้างโมเดลจำแนกประเภท	7
2.3 ขั้นตอนการใช้งานโมเดลจำแนกประเภท	7
2.4 วิธีการหลักของตัวจำแนกประเภทแบบเบย์ (Naive Bayes Classifier).....	12
2.5 ทฤษฎีที่เกี่ยวข้อง.....	14
3.1 กระบวนการสร้างโมเดลการค้นหาโรงแรม	16
3.2 ขั้นตอนการเตรียมข้อมูล	17
3.3 ตัวอย่างข้อมูลโรงแรม	19
3.4 ความคิดเห็นของผู้ใช้บริการโรงแรม	20
3.5 ขั้นตอนการสร้างโมเดล.....	22
3.6 ขั้นตอนการใช้งานโมเดล.....	23
3.7 แผนภาพการทำงานของผู้ใช้งาน.....	25
3.8 ผังงานการทำงานของระบบ	26
3.9 ผังงานแสดงการทำงานของเอเจนต์ ที่ฝั่งเซิร์ฟเวอร์.....	27
3.10 แผนภาพแสดงการดูรายละเอียดโรงแรม	28
3.11 แผนภาพแสดงการดูลำดับโรงแรม	28
3.12 แผนภาพแสดงการค้นหาโรงแรม	29
3.13 แบบจำลองความสัมพันธ์เอนทิตีของระบบอัจฉริยะในการค้นหาโรงแรม	30
3.14 หน้าหลักของระบบ.....	31
3.15 หน้าของการค้นหาโรงแรม	32
3.16 หน้าแสดงรายละเอียดของโรงแรม.....	32
3.17 หน้าแสดงรายละเอียดของโรงแรม.....	33
3.18 หน้าแสดงรูปภาพเพิ่มเติมของโรงแรม.....	33
4.1 กระบวนการทดลอง.....	34
4.2 หน้าหลักของระบบอัจฉริยะในการค้นหาโรงแรม	37
4.3 หน้าจอแสดงผลการค้นหา	38
4.4 หน้าจอแสดงผลการเรียงลำดับด้านห้องพัก 5 โรงแรมแรก.....	39
4.5 หน้าจอแสดงผลการเรียงลำดับด้านห้องพักทั้งหมด 60 โรงแรม	39

สารบัญรูป (ต่อ)

รูปที่	หน้า
4.6 หน้าจอแสดงรายละเอียดของโรงแรม	40
4.7 หน้าจอแสดงรายละเอียดแบบเต็มของโรงแรม	40
4.8 หน้าจอแสดงรูปภาพของโรงแรม.....	41
ก.1 เว็บไซต์ www.apachefriends.org	46
ก.2 หน้า phpMyadmin	46
ก.3 หน้าสร้างฐานข้อมูล.....	47
ก.4 หน้า Import ไฟล์ SQL.....	47
ก.5 หน้าเว็บเพจระบบอัจฉริยะในการค้นหาโรงแรม.....	48
ข.1 การดึงไอทีของความคิดเห็น.....	50
ข.2 การดึงคะแนนความคิดเห็นโดยรวมของความคิดเห็น.....	51
ข.3 การดึงคะแนนความคิดเห็นในแต่ละด้านของความคิดเห็น	51
ข.4 การดึงความคิดเห็น.....	52
ข.5 การกำหนดสถานะ.....	52

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญ

ในปัจจุบันเว็บแอปพลิเคชันยังคงได้รับความนิยมอย่างต่อเนื่องและมีการใช้งานอย่างแพร่หลายเนื่องจากมีความสะดวกในการใช้งาน และด้วยความก้าวหน้าของเทคโนโลยีทำให้ smartphone สามารถเชื่อมต่อกับอินเทอร์เน็ต ส่งผลให้สามารถเข้าถึงเว็บแอปพลิเคชันได้ทุกที่ทุกเวลา เห็นได้ว่าเว็บแอปพลิเคชันยังคงมีบทบาทและตอบสนองความต้องการของผู้ใช้สูงขึ้นเรื่อยๆ ทำให้เว็บแอปพลิเคชันเป็นส่วนหนึ่งของชีวิตประจำวัน

สืบเนื่องจากปัจจุบันอุตสาหกรรมการท่องเที่ยวในประเทศไทยได้เจริญเติบโตส่งผลให้จำนวนโรงแรมที่พักเพิ่มขึ้นมากมาย เกิดการแข่งขันเพื่อดึงดูดนักท่องเที่ยว และพบว่าเว็บไซต์ที่รวบรวมข้อมูลห้องพักบางเว็บไซต์เปิดให้ผู้ใช้มีส่วนร่วมในการแสดงความคิดเห็น สอดคล้องกับพฤติกรรมของผู้ใช้ที่เปลี่ยนแปลงไปทำให้สื่อออนไลน์เข้ามามีอิทธิพลต่อการตัดสินใจซื้อสินค้าและบริการมากขึ้น ซึ่งข้อมูลเหล่านั้นสามารถใช้วัดคุณภาพได้ในระดับหนึ่ง โดยมีแนวคิดที่จะนำข้อมูลความคิดเห็นต่างๆ เหล่านี้มาใช้ในการค้นหาข้อมูลที่พัก เพื่อให้เกิดความพึงพอใจและตรงความต้องการของผู้ใช้มากที่สุด

จากการศึกษาข้อมูลพบว่าเว็บไซต์ให้บริการค้นหาที่พัก เช่น เว็บไซต์ Agoda (www.agoda.com) เว็บไซต์ Expedia (www.expedia.co.th) หรือ เว็บไซต์ Tripadvisor (www.tripadvisor.com) เป็นต้น ซึ่งเว็บไซต์เหล่านี้บางเว็บไซต์ไม่สามารถเลือกโรงแรมที่มีบริการหรือสภาพแวดล้อมตามที่ต้องการ บางเว็บไซต์ไม่สามารถจัดเรียงคุณภาพด้านต่างๆ ของโรงแรมได้ และอาจได้ข้อมูลที่เกิดจากเจ้าของโรงแรมเข้ามาโหวตเอง ด้วยเหตุนี้จึงได้พัฒนาเว็บแอปพลิเคชันด้วยเทคนิคตัวจำแนกประเภทแบบเบย์อย่างง่าย โดยเว็บแอปพลิเคชันนี้สามารถตอบโจทย์ทั้งการเลือกบริการและสิ่งแวดล้อมที่โรงแรมมีตามที่ต้องการ หรือจัดเรียงโรงแรมตามคุณภาพด้านต่างๆ ที่เชื่อถือได้ในระดับหนึ่งเพราะมาจากความคิดเห็นของผู้ใช้งานจริง

1.2 วัตถุประสงค์ของปัญหาพิเศษ

1. เพื่อศึกษาการพัฒนาเว็บแอปพลิเคชันด้วยโปรแกรมประยุกต์ weka และภาษา PHP อีกทั้งยังเป็นการศึกษาการบริหารจัดการฐานข้อมูลด้วย PHP My Admin บนฐานข้อมูล MySQL
2. พัฒนาระบบการค้นหาโรงแรมโดยใช้เทคนิคตัวจำแนกประเภทแบบเบย์อย่างง่าย
3. เพื่ออำนวยความสะดวกสำหรับผู้ที่ต้องการค้นหาโรงแรมผ่านทางระบบอินเทอร์เน็ต

1.3 ขอบจำกัดและขอบเขตของปัญหาพิเศษ

1. ข้อมูลที่ใช้สำหรับการทดลอง คือ ข้อมูลความคิดเห็นของโรงแรมต่างๆ โดยรวบรวมจากเว็บไซต์ www.tripadvisor.com
2. ข้อมูลโรงแรมที่ใช้ในการพัฒนาเว็บแอปพลิเคชันมีจำนวน 60 โรงแรมในเขตการปกครองพิเศษพัทยา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ด้านผู้พัฒนาเว็บแอปพลิเคชัน
 - เรียนรู้และเข้าใจการวิเคราะห์ และออกแบบเว็บแอปพลิเคชัน
 - เรียนรู้และเข้าใจในการติดตั้ง ใช้งานเครื่องมือ สามารถสร้างและพัฒนาเว็บแอปพลิเคชันได้
 - เรียนรู้และเข้าใจการพัฒนาเว็บแอปพลิเคชันโดยใช้ภาษา PHP, HTML, CSS และ java script
 - เรียนรู้และเข้าใจการทำเหมืองข้อมูลโดยการนำเทคนิคตัวจำแนกประเภทแบบเบย์อย่างง่าย และโปรแกรม weka มาประยุกต์ใช้ในการค้นหา
2. ด้านผู้ใช้
 - สามารถค้นหาโรงแรมที่ตรงกับความต้องการของผู้ใช้ โดยการค้นหาแบบพิมพ์ข้อความ
 - ทราบถึงข้อมูลต่างๆของโรงแรมที่มีอยู่บนเว็บแอปพลิเคชัน เช่น สถานที่ตั้ง รูปภาพ

1.5 ขั้นตอนในการดำเนินการ

1. เก็บรวบรวมข้อมูลตัวอย่างความคิดเห็นจากเว็บไซต์ www.tripadvisor.com
2. ศึกษาทฤษฎีการจำแนกแบบเบย์อย่างง่าย
3. ศึกษาโปรแกรมประยุกต์ weka เพื่อใช้ประมวลผลในการเตรียมข้อมูล
4. ศึกษาและเรียนรู้เครื่องมือที่ใช้พัฒนาเว็บแอปพลิเคชัน
5. ทำการพัฒนาเว็บแอปพลิเคชันที่ออกแบบไว้
6. ทดสอบกระบวนการทำงานทั้งหมดของเว็บแอปพลิเคชันให้สามารถทำงานได้ตามที่ถูกต้อง ออกแบบไว้ และปรับปรุงให้เสร็จสมบูรณ์

1.6 อุปกรณ์ที่ใช้ในการทำปัญหาพิเศษ

1. ฮาร์ดแวร์
 - โน้ตบุ๊ก (Notebook)
 - หน่วยประมวลผล(CPU) Intel Core i3 - 2370M 2.4 GHz
 - ฮาร์ดดิสก์ 500 GB
 - RAM 8 GB
2. ซอฟต์แวร์
 - ระบบปฏิบัติการ Windows 8
 - ระบบการบริหารการบริการด้านเว็บ xampp v5.6.11(Apache 2.4.12, MySQL 5.6.25, PHP 5.6.11, phpMyadmin 4.4.12)
 - โปรแกรมเพื่อใช้ในการพัฒนา คือ Adobe Dreamweaver CS6
 - โปรแกรมเพื่อใช้ในการจัดการฐานข้อมูล คือ MySQL workbench
 - โปรแกรมเพื่อใช้ในการออกแบบกราฟิก คือ Adobe Photoshop CS6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

เนื้อหาบทนี้ประกอบด้วยทฤษฎีและงานวิจัยที่เกี่ยวข้อง ซึ่งกล่าวถึงความรู้เบื้องต้นเกี่ยวกับการทำเหมืองข้อมูล ประกอบไปด้วย 5 ส่วนคือ ความหมายของการทำเหมืองข้อมูล ขั้นตอนการทำเหมืองข้อมูล เทคนิคการทำเหมืองข้อมูล ประโยชน์ของการทำเหมืองข้อมูล และแนวโน้มการประยุกต์ใช้งานเหมืองข้อมูล นอกจากนี้ยังกล่าวถึงทฤษฎีที่เกี่ยวข้องกับการทำปัญหาพิเศษนี้ ได้แก่ การจำแนกประเภทข้อมูล การจำแนกประเภทแบบเบย์อย่างง่าย เทคนิคการคัดเลือกคุณลักษณะ มาตราวัดความสอดคล้องของเคนดอลล์ และจะกล่าวถึงงานวิจัยต่างๆที่เกี่ยวข้อง โดยแบ่งประเภทงานวิจัยที่เกี่ยวข้องเป็น 2 ส่วนคือ การวิเคราะห์ความรู้สึกของสถานะบนเฟสบุ๊คด้วยตัวจำแนกประเภทแบบเบย์ และการจำแนกแบบเบย์เป็นการจำแนกประเภทเอกสารที่ดีที่สุด

2.1 การทำเหมืองข้อมูล

การทำเหมืองข้อมูล เป็นการค้นหาความรู้ในฐานข้อมูล (Knowledge Discovery in Database - KDD) จากข้อมูลจำนวนมาก เพื่อหารูปแบบ และความสัมพันธ์ที่อยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์

2.1.1 ความหมายของการทำเหมืองข้อมูล

การทำเหมืองข้อมูล (Data Mining)[1] เป็นขั้นตอนการวิเคราะห์ข้อมูล เพื่อแยกประเภท จำแนกรูปแบบและความสัมพันธ์ของข้อมูลจากฐานข้อมูลขนาดใหญ่หรือคลังข้อมูล และนำเสนอสารสนเทศที่ได้ไปใช้ในการตัดสินใจ โดยใช้เทคนิคบางประการเพื่อค้นหาข้อมูลที่ถูกซ่อนอยู่ จำแนก และเชื่อมโยงความสัมพันธ์ของข้อมูลเหล่านั้นจากคลังข้อมูล ทำให้ค้นพบรูปแบบความสัมพันธ์ที่ไม่เคยมีมาก่อนจนกลายเป็นการค้นหาองค์ความรู้ใหม่ (Knowledge Discovery) ซึ่งเป็นวัตถุประสงค์หลักของการทำเหมืองข้อมูล องค์ความรู้ใหม่ที่ได้อาจรวมถึงกฎ (Rules) บางอย่างเพื่อใช้เป็นแนวทางในการตัดสินใจ และประเมินผลลัพธ์การตัดสินใจได้

2.1.2 ขั้นตอนการทำเหมืองข้อมูล มีขั้นตอนหลักอยู่ 3 ขั้นตอน คือ

1. Data Preprocessing คือขั้นตอนการจัดเตรียมข้อมูล เพื่อให้มั่นใจว่าคุณภาพของข้อมูลที่ถูกเลือกนั้นเหมาะสม จะต้องทำการคัดข้อมูลที่ไม่เกี่ยวข้องหรือข้อมูลรบกวน (Noise Data) ออกจากข้อมูลดิบ สามารถแบ่งเป็นขั้นตอนย่อย ดังนี้
 - Data cleaning เป็นขั้นตอนในการกำจัดข้อมูลที่ไม่ต้องการ หรือข้อมูลที่ไม่เป็นประโยชน์ต่อการใช้งาน หรือข้อมูลที่มีความผิดพลาด
 - Data integration เป็นขั้นตอนในการรวบรวมข้อมูลทั้งหมดที่มีจากแหล่งข้อมูลต่างๆ มาไว้ด้วยกัน

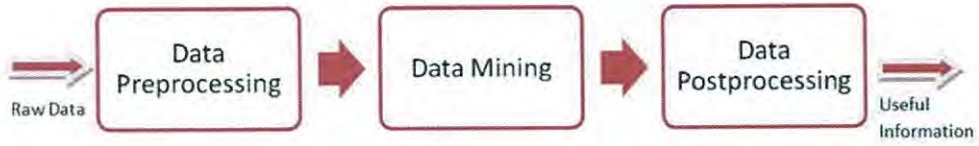
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Data selection เป็นการคัดเลือกข้อมูลที่เกี่ยวข้องกับข้อมูลที่จะวิเคราะห์ เลือกข้อมูลที่มีความสัมพันธ์กัน ส่งผลต่อกัน และเป็นประโยชน์ต่อการทำนาย
- Data transformation เป็นการจัดรูปแบบข้อมูลที่ได้จากขั้นตอนการคัดเลือกข้อมูล ให้มีความเหมาะสมต่อการทำนาย เช่น การจัดระเบียบข้อมูลที่สัมพันธ์กันมาไว้ในระเบียนชุดเดียวกัน หรือการแปลงค่าตัวเลขให้อยู่ในช่วงที่กำหนด (Normalization) เป็นต้น
- Data Reduction เป็นการทำให้ข้อมูลตั้งต้นมีขนาดลดลง และสามารถนำข้อมูลนี้ไปใช้สร้างโมเดล แล้วประสิทธิภาพในการทำงานของโมเดลไม่ลดลง เนื่องจากข้อมูลแต่ละตัวจะมีความสำคัญต่อการจัดกลุ่มข้อมูลไม่เท่ากัน ด้วยเทคนิคการเลือกข้อมูลที่ดีจะทำให้สามารถเลือกข้อมูลที่มีความสำคัญ และสามารถใช้เป็นตัวแทนของข้อมูลส่วนใหญ่ได้

ซึ่งในขั้นตอนนี้จะเป็นขั้นตอนที่ใช้เวลาในการดำเนินงานนานที่สุด หากจัดรูปแบบข้อมูลได้ไม่เหมาะสมจะส่งผลให้โมเดลที่สร้างไม่มีความแม่นยำในการทำนาย การทราบถึงปัจจัยที่ส่งผลต่อการทำนายจะทำให้วิเคราะห์ข้อมูลได้ถูกต้องและการทำนายจะมีความแม่นยำ

2. Data Mining คือขั้นตอนการนำข้อมูลที่คัดเลือกแล้วมาสร้างโมเดล โดยจะต้องทำการเลือกเทคนิคที่เหมาะสมกับรูปแบบชุดข้อมูล หลังจากได้เทคนิคที่เหมาะสมแล้ว จะทำการสอน (Train) ให้โมเดลเรียนรู้ลักษณะของข้อมูลว่า ชุดข้อมูลทั้งหมดมีความสัมพันธ์กันอย่างไร และทิศทางในการวิเคราะห์เป็นอย่างไร โดยในการสอนให้โมเดลเรียนรู้ นั้นจำเป็นต้องมีการกำหนดพารามิเตอร์ (Parameter) หรือค่าตัวแปรต่างๆ ให้เหมาะสม ซึ่งในการพิจารณาค่าพารามิเตอร์นั้นขึ้นอยู่กับเทคนิคที่เลือกใช้ และประสบการณ์ หลังจากนั้นจึงนำโมเดลที่ได้ไปทดสอบหาค่าความผิดพลาดของโมเดล จากการนำข้อมูลจริงที่เตรียมไว้สำหรับการทดลองมาป้อนลงในโมเดลแล้วดูผลของการทำนายที่ได้
3. Data Postprocessing คือขั้นตอนสุดท้ายของการทำเหมืองข้อมูล เป็นขั้นตอนการประเมินผลและนำสิ่งที่ได้มานำเสนอในรูปแบบของการใช้งาน หากผลการทดสอบ (Test) ไม่เป็นที่น่าพอใจ จะต้องทำการเตรียมข้อมูลใหม่ เพื่อให้ได้ค่าความถูกต้อง (Accuracy) มากที่สุด หรือค่าความผิดพลาด (Error) ในการทำนายน้อยที่สุด หากความผิดพลาดยังมีอยู่ อาจจำเป็นต้องเลือกเทคนิคในการทำเหมืองข้อมูลใหม่ ขึ้นอยู่กับปัญหาที่เกิดขึ้น และในขั้นตอนแบ่งออกเป็น 2 ขั้นตอนย่อยๆ ดังนี้
 - Pattern evaluation เป็นการประเมินผลของโมเดลที่ได้ว่า เกิดความผิดพลาดมากน้อยเพียงใด การทำนายที่ได้นั้นเป็นไปตามความต้องการหรือไม่
 - Knowledge presentation เป็นการแสดงผลของการทำงานที่ได้จากโมเดล โดยทำเป็นรูปแบบที่สะดวกต่อการใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.1 ขั้นตอนการทำเหมืองข้อมูล

2.1.3 เทคนิคการทำเหมืองข้อมูล

การทำเหมืองข้อมูลเป็นเทคนิคในการค้นหาความรู้จากฐานข้อมูลความรู้ขนาดใหญ่ จึงรวมเอาศาสตร์หลายแขนงมารวมไว้ด้วยกัน เช่น สถิติ (Statistics) วิทยาศาสตร์สารสนเทศ (Information science) เป็นต้น ซึ่งจะทำให้เกิดกระบวนการค้นหาความรู้ในแบบต่างๆ โดยรูปแบบการค้นหาความรู้มีดังนี้

1. การจำแนกประเภท (Classification) เป็นกระบวนการสร้างโมเดลจัดการข้อมูลที่มีความใกล้เคียงกัน เพื่อใช้ในการทำนายชุดข้อมูลว่า อยู่ในประเภทใดของชุดข้อมูลที่ได้ทำการแบ่งไว้แล้ว ซึ่งชุดข้อมูลที่แบ่งไว้เกิดจากชุดข้อมูลที่มีอยู่แล้ว (Training data) ตัวอย่างเทคนิคการจำแนกประเภท ได้แก่ ต้นไม้ตัดสินใจ (Decision tree) และโครงข่ายประสาทเทียม (Neural Network) เป็นต้น
2. การวิเคราะห์เพื่อจัดกลุ่ม (Clustering Analysis) ใช้ในการแบ่งกลุ่มตัวอย่างตัวอย่างที่อยู่ในกลุ่มเดียวกันจะต้องมีความสัมพันธ์กันมากกว่าตัวอย่างที่อยู่ต่างกลุ่มกัน ตัวอย่างที่อยู่ต่างกลุ่มกันมีความสัมพันธ์กันน้อยหรือไม่มีความสัมพันธ์กันเลย ตัวอย่างการวิเคราะห์เพื่อจัดกลุ่ม ได้แก่ อัลกอริทึม K-mean Clustering เป็นต้น
3. การวิเคราะห์กฎความสัมพันธ์ (Association rule Discovery) เป็นการค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่เพื่อนำไปใช้ในการวิเคราะห์หรือทำนายปรากฏการณ์ต่างๆ หรือมาจากการวิเคราะห์การซื้อสินค้าของลูกค้า เรียกว่า “ Market Basket Analysis ” ซึ่งประเมินจากข้อมูลในตารางที่รวบรวมไว้ ผลการวิเคราะห์ที่ได้จะเป็นคำตอบของปัญหาซึ่งการวิเคราะห์แบบนี้เป็นการใช้ “กฎความสัมพันธ์” (Association Rule) เพื่อหาความสัมพันธ์ของข้อมูล

2.1.4 ประโยชน์ของการใช้เหมืองข้อมูล

เหมืองข้อมูลถูกนำมาใช้สนับสนุนการตัดสินใจ โดยการสร้างมูลค่าเพิ่มให้กับข้อมูลที่มีอยู่ ประโยชน์ที่ได้ สรุปได้ดังนี้

- การเอาชนะคู่แข่ง บริษัทต่างๆพยายามแย่งส่วนแบ่งจากการตลาดที่สามารถสร้างผลกำไรให้กับบริษัทได้ โดยนำเหมืองข้อมูลมาใช้ในการแย่งส่วนแบ่งทางการตลาด และป้องกันไม่ให้เกิดการเสียส่วนแบ่งทางการตลาด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ให้เกิดความรู้ที่สามารถนำมาใช้หรือประกอบการตัดสินใจได้ เนื่องจากใช้เทคนิคที่ซับซ้อนในการสร้างโมเดลที่ใช้ข้อมูลซึ่งรวบรวมมาจากแหล่งต่างๆ ความรู้ที่ได้จึงช่วยให้องค์กรสามารถคาดการณ์อนาคต และสามารถเจาะกลุ่มตลาดได้ถูกต้องมากขึ้น
- กำหนดเป้าหมายกลุ่มลูกค้าได้อย่างมีประสิทธิภาพมากขึ้น ทำให้ลดโอกาสของความเสียหายต่างๆ และมียอดขายเพิ่มขึ้น

2.1.5 แนวโน้มและการประยุกต์ใช้งานเหมืองข้อมูล

เนื่องจากในปัจจุบันมีการนำหลักการของเหมืองข้อมูลมาใช้อย่างแพร่หลาย จึงมีการค้นคว้าและพัฒนาเพื่อประยุกต์ใช้กับงานในหลายๆด้าน ตัวอย่างการประยุกต์ใช้งานที่น่าสนใจในปัจจุบัน ได้แก่

- ด้านการแพทย์ (Biomedical and DNA Data Analysis) ส่วนมากนำไปใช้ในการวิเคราะห์รูปแบบการจัดเรียงตัวของหน่วยพันธุกรรมเพื่อหาสาเหตุของความผิดปกติที่ทำให้เกิดโรครวมถึงการใช้ในการวินิจฉัยโรค การป้องกัน และการรักษาโรค
- ด้านการเงิน (Financial Analysis) เกี่ยวกับบริษัทเงินทุนหรือธนาคารต่างๆ เช่น การวิเคราะห์การให้สินเชื่อ การแบ่งกลุ่มลูกค้าเพื่อหาเป้าหมายทางการตลาด เป็นต้น
- ด้านการขาย (Retail Industry) ใช้หากกลยุทธ์ที่ทำให้เกิดการได้เปรียบคู่แข่งทางการค้า เช่น การหาลักษณะการซื้อของลูกค้า ความสัมพันธ์ของการซื้อกับช่วงเวลา เป็นต้น ซึ่งช่วยให้สามารถหาวิธีตอบสนองลูกค้าได้มากที่สุด
- ด้านโทรคมนาคม (Telecommunication Industry) สนับสนุนการให้บริการด้านการติดต่อสื่อสาร เช่น การหาความสัมพันธ์ของการใช้บริการกับช่วงเวลา เป็นต้น

2.2 การจำแนกประเภทข้อมูล

การจำแนกประเภทข้อมูลคือกระบวนการสร้างโมเดลจำแนกประเภทข้อมูล (Classification) เพื่อทำนายข้อมูลที่จะเกิดในอนาคต โดยสร้างจากการหาความสัมพันธ์ของข้อมูล ซึ่งแบ่งข้อมูลเป็นสองส่วน ส่วนแรกคือข้อมูลที่ใช้ฝึกสอน (Training Data) เป็นชุดข้อมูลที่ใช้สร้างโมเดลจำแนกประเภทข้อมูล ส่วนที่สองคือข้อมูลสำหรับทดสอบ (Testing Data) เป็นชุดข้อมูลเพื่อทดสอบประสิทธิภาพของโมเดลการจำแนกประเภทข้อมูล โดยนำโมเดลที่สร้างมาจำแนกข้อมูลที่ใช้สำหรับทดสอบ สำหรับการสร้างโมเดลจำแนกประเภทข้อมูลแบ่งออกเป็น 2 ขั้นตอนมีดังนี้

1. Model Construction เป็นขั้นตอนการสร้างโมเดลจำแนกประเภท โดยอาศัยการเรียนรู้จากข้อมูลที่ได้มีการกำหนดคลาสหรือเรียกว่าข้อมูลสำหรับฝึกสอน ซึ่งสามารถสร้างโมเดลจำแนกประเภทได้ด้วยอัลกอริทึมทางการเรียนรู้ด้วยเครื่อง (Machine Learning) ตัวอย่างอัลกอริทึมการเรียนรู้ของเครื่องที่ใช้สร้างโมเดลได้แก่ ต้นไม้ตัดสินใจ (Decision Tree) ต้นไม้จำแนกประเภทแบบเบย์อย่างง่าย (Naive Bayes Classifier) และโครงข่ายประสาทเทียม (Artificial Neural Networks : ANNs)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.2 ขั้นตอนการสร้างโมเดลจำแนกประเภท

2. Model Usage เป็นขั้นตอนการนำโมเดลจำแนกประเภทที่สร้างมาใช้กับข้อมูลที่ไม่เคยพบมาก่อน (Unseen data) .เพื่อทำนายและจำแนกประเภทให้กับข้อมูลนั้น และวัดประสิทธิภาพในการทำนายด้วยข้อมูลทดสอบซึ่งเป็นข้อมูลคนละส่วนกับข้อมูลสำหรับฝึกสอน เมื่อโมเดลที่สร้างขึ้นมีประสิทธิภาพในการทำนายข้อมูลในระดับที่ยอมรับได้ จึงนำโมเดลนั้นมาใช้ในการจำแนกประเภทข้อมูลที่ไม่เคยพบมาก่อน



รูปที่ 2.3 ขั้นตอนการใช้งานโมเดลจำแนกประเภท

2.3 การจำแนกประเภทแบบเบย์อย่างง่าย

การจำแนกประเภทแบบเบย์อย่างง่าย (Naive Bayes Classifier) เป็นวิธีการจำแนกประเภทข้อมูลที่มีประสิทธิภาพวิธีหนึ่งที่อยู่อาศัยหลักการความน่าจะเป็น (Probability) และคำนวณการแจกแจงความน่าจะเป็นตามสมมติฐานที่ตั้งให้กับข้อมูล โดยการเรียนรู้ปัญหาที่เกิดขึ้น เหมาะกับกรณีของเซตตัวอย่างที่มีจำนวนและคุณสมบัติ (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน เพื่อนำมาสร้างการจำแนกเงื่อนไขการจำแนกข้อมูลใหม่ จากการคำนวณ ตัวอย่างใหม่ที่ได้จะถูกนำมาปรับเปลี่ยนการแจกแจงซึ่งมีผลต่อการเพิ่มหรือลดความน่าจะเป็นของข้อมูล ข้อมูลใหม่ที่เกิดขึ้นและรูปแบบที่ตั้งไว้ให้กับข้อมูลจะถูกปรับเปลี่ยนไปตามข้อมูลใหม่ โดยรวมกับข้อมูลเดิมที่มี

กำหนดให้เป็นสัญลักษณ์ทางคณิตศาสตร์ คือ $P(A|B)$ โดยมีสูตรการคำนวณความน่าจะเป็นแบบมีเงื่อนไขดังสมการที่ 2.1

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (2.1)$$

โดยที่ $P(A|B)$ คือ ค่าความน่าจะเป็นที่เกิดเหตุการณ์ A เมื่อเกิดเหตุการณ์ B ก่อนหน้า
 $P(A)$ คือ ค่าความน่าจะเป็นที่เกิดเหตุการณ์ A
 $P(B)$ คือ ค่าความน่าจะเป็นที่เกิดเหตุการณ์ B

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กำหนดให้ความน่าจะเป็นของข้อมูลที่จะเป็นประเภท v_j สำหรับข้อมูลที่มีคุณลักษณะ n ตัว $X = \{a_1, a_2, \dots, a_n\}$ หรือใช้สัญลักษณ์ว่า $P(a_1, a_2, \dots, a_n | v_j)$ ดังสมการที่ 2.2

$$P(a_1, a_2, \dots, a_n | v_j) = P(v_j) \prod_{i=1}^n P(a_i | v_j) \quad (2.2)$$

โดยที่ $P(v_j)$ คือ ค่าความน่าจะเป็นของข้อมูลที่ให้คลาส v_j

$P(a_i | v_j)$ คือ ค่าความน่าจะเป็นของข้อมูลคุณลักษณะที่ i มีค่า a_i และให้ประเภท v_j

$P(a_1, a_2, \dots, a_n | v_j)$ คือ ค่าความน่าจะเป็นของข้อมูลทดสอบ (ที่มีค่าคุณลักษณะ a_1, a_2, \dots, a_n) ที่จะให้ประเภท v_j

\prod คือ ผลคูณของค่า $P(a_i | v_j)$ ทั้งหมด $i = 1, 2, 3, \dots, n$ และ $j = 1, 2, 3, \dots, n$

ขั้นตอนการนำการจำแนกประเภทแบบเบย์อย่างง่ายไปทำนายข้อมูลหรือจำแนกประเภทข้อมูล มีดังต่อไปนี้

1. หาค่าความน่าจะเป็นของคุณลักษณะที่พบในแต่ละกลุ่มโดยนำค่า $P(a_1, a_2, \dots, a_n | v_j)$ จากสมการที่ 2.2 มาคูณกับค่าความน่าจะเป็นของกลุ่มนั้นๆ คือ $P(v_j)$ ได้เท่ากับ v_{NB}
2. นำค่าที่ได้มาเปรียบเทียบกัน กลุ่มที่มีค่าความน่าจะเป็นสูงสุด คือ คำตอบ ดังนั้นจะได้สูตรการคำนวณการจำแนกประเภทแบบเบย์อย่างง่าย ดังสมการที่ 2.3

$$v_{NB} = \operatorname{argmax}_{v_j \in V} P(v_j) \times \prod_{i=1}^n P(a_i | v_j) \quad (2.3)$$

การจัดกลุ่มเอกสารเป็นเทคนิคสำคัญของการจำแนกประเภทข้อความ เพื่อจำแนกเอกสารโดยอัตโนมัติ มีการนำการจำแนกประเภทแบบเบย์อย่างง่ายไปประยุกต์ใช้งานในด้านการจำแนกประเภทข้อความ (Text Classification) การวินิจฉัย (Diagnosis) และพบว่าใช้งานได้ดีไม่ต่างจากการจำแนกประเภทวิธีอื่น

สมมติให้เอกสารหนึ่งๆ คือ ตัวอย่างหนึ่งตัว โดยใช้คำที่ปรากฏในเอกสารเป็นคุณลักษณะของเอกสาร ให้ a_1 คือ คุณลักษณะของคำที่หนึ่ง a_2 คือ คุณลักษณะของคำที่สองตามลำดับ จากนั้นสร้างการเรียนรู้ข้อมูลโดยใช้ตัวอย่างสอน เพื่อประมาณค่าความน่าจะเป็นจากสมมติฐานเรื่องความไม่ขึ้นต่อกันของคุณลักษณะของการจำแนกแบบเบย์ กำหนดสมการหาค่าความน่าจะเป็นของเอกสาร ดังสมการที่ 2.4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$P(doc | v_j) = \prod_{i=1}^{length(doc)} P(a_i = w_k | v_j) \quad (2.4)$$

โดยที่ a_i คือ คุณลักษณะตัวที่ i

w_k คือ คำในพจนานุกรมคำที่ k

$P(a_i = w_k | v_j)$ คือ ความน่าจะเป็นที่คำในตำแหน่งที่ i ของคำที่ k ในรายการของคำที่พบในประเภทเอกสาร v_j

$P(doc | v_j)$ คือ ความน่าจะเป็นของแต่ละประเภทเอกสาร v_j

สรุปวิธีการจำแนกประเภทเอกสารโดยการใช้การจำแนกแบบเบย์[2] ได้ดังนี้

1. นำคำที่ผ่านการคัดเลือกจากเอกสารมาสร้างเป็นข้อมูลตัวอย่าง
2. คำนวณหาค่าความน่าจะเป็นของประเภทเอกสาร $P(v_j)$ และคำนวณหาค่าความน่าจะเป็นของคำที่พบทั้งหมดในประเภทเอกสาร $P(a_i | v_j)$ ซึ่งคำนวณได้จาก
3. คำนวณหาค่า $P(v_j)$ ได้จากสมการที่ 2.5

$$P(v_j) = \frac{\text{จำนวนเอกสารในประเภทเอกสารนั้นๆ}}{\text{จำนวนเอกสารทั้งหมด}} \quad (2.5)$$

4. คำนวณหาค่า $P(a_i | v_j)$ ได้จากสมการที่ 2.6

$$P(a_i | v_j) = \frac{\text{ผลรวมความถี่ของคำที่พบในประเภทเอกสารนั้นๆ}}{\text{จำนวนเอกสารในประเภทเอกสารนั้นๆ}} \quad (2.6)$$

5. คำนวณหาค่า $P(v_j) \times \prod_{i=1}^n P(a_i | v_j)$ หมายถึง ค่าความน่าจะเป็นของเอกสารในแต่ละประเภทเอกสาร คือ ผลคูณระหว่างค่าความน่าจะเป็นของกลุ่มเอกสารกับค่าความน่าจะเป็นของคำที่พบทั้งหมดในเอกสาร
6. เลือกประเภทเอกสารที่มีค่าความน่าจะเป็นมากที่สุด เป็นคำตอบ

2.4 เทคนิคการคัดเลือกคุณลักษณะ

เทคนิคการคัดเลือกคุณลักษณะ[3] เป็นการลดจำนวนคุณลักษณะที่ไม่เกี่ยวข้องออก ทำให้เหลือเฉพาะคุณลักษณะที่มีความสัมพันธ์กันเท่านั้น ข้อดีของการลดคุณลักษณะ คือ นำคุณลักษณะที่กรองแล้วมาสอนทำให้ผลการจำแนก (Classification) ได้ค่าความถูกต้องสูงขึ้นและมีเวลาในการประมวลผลลดลง โดยประยุกต์ใช้เทคนิคการคัดเลือกคุณลักษณะค่าสารสนเทศเกน (Information Gain)

2.4.1 ค่าสารสนเทศเกน

ค่าสารสนเทศเกนใช้ในการเลือกคุณลักษณะของข้อมูล ซึ่งถ้าคุณลักษณะใดมีค่าสารสนเทศเกนสูงแสดงว่า คุณลักษณะนั้นมีความสำคัญในการจำแนกประเภทข้อมูล โดยวิธีการหาค่าสารสนเทศเกนนั้นพิจารณาจากการวัดค่าความไร้ระเบียบ (Entropy) เพื่อคัดเลือกคุณลักษณะที่ให้ค่าสารสนเทศเกนสูงที่สุด โดยค่าความไร้ระเบียบสามารถหาได้จากสมการที่ 2.7

$$Entropy(S) = \sum_{i=1}^c -P_i \log_2 P_i \quad (2.7)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่ P_i คือ สัดส่วนของข้อมูลที่อยู่ในคลาส i เทียบกับข้อมูลทั้งหมด
 C คือ จำนวนคลาสทั้งหมด

ค่าเอนโทรปี คือ ค่าที่บ่งบอกถึงควมไร้ระเบียบของข้อมูล ถ้าข้อมูลที่มีค่าความไร้ระเบียบสูง ค่าเอนโทรปีจะมีค่าสูง และในทางกลับกันถ้าข้อมูลมีความไร้ระเบียบต่ำ ค่าเอนโทรปีจะมีค่าน้อย ซึ่งเมื่อนำมาใช้กับค่าสารสนเทศเกณฑ์จะได้ดังสมการที่ 2.8

$$Gain(S, A) = Entropy(S) - \sum_{v \in value(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2.8)$$

โดยที่ S คือ ข้อมูลก่อนที่จะใช้คุณลักษณะ A เป็นตัวแบ่ง
 $|S_v|$ คือ จำนวนข้อมูลที่มีค่าของคุณลักษณะ A เท่ากับ v
 $|S|$ คือ จำนวนข้อมูลทั้งหมดในชุดข้อมูล S
 $Entropy(S)$ คือ ค่าความไร้ระเบียบของชุดข้อมูล S
 $Entropy(S_v)$ คือ ค่าความไร้ระเบียบของข้อมูลที่มีค่าคุณลักษณะ A เท่ากับ v
 $Value(A)$ คือ เซตของค่าของคุณลักษณะ A ที่เป็นไปได้

อัลกอริทึมจะคำนวณค่าสารสนเทศเกณฑ์ของแต่ละคุณลักษณะ ซึ่งคุณลักษณะที่มีค่าสารสนเทศเกณฑ์สูงจะถูกเลือกให้เป็นคุณลักษณะที่จะนำไปใช้ต่อไป

2.5 มาตรการวัดความสอดคล้องของเคนดอลล์

มาตรการวัดความสอดคล้องของเคนดอลล์ (Kendall Tau Correlation)[4] ใช้เพื่อวัดค่าความสอดคล้องระหว่างผลลัพธ์ที่ได้จากการทำนายกับผลลัพธ์ที่แท้จริงโดยจะใช้การจับคู่ระหว่างอันดับของผลลัพธ์ที่ได้จากโมเดลกับผลลัพธ์ที่แท้จริงมีคู่อันดับ (i, j) โดยที่อันดับ i และ j มีค่าที่ไม่ซ้ำกัน เมื่อ i มีค่าอยู่ในช่วง $1 \leq i \leq n$ j มีค่าอยู่ในช่วง $1 \leq j \leq n$ และ n คือจำนวนอันดับทั้งหมด คู่อันดับ (i, j) จะมีความสอดคล้อง (concordant) กันเมื่อ (x_i, y_i) และ (x_j, y_j) เป็นไปในทิศทางเดียวกันคือ $x_i > x_j, y_i > y_j$ หรือ $x_i < x_j, y_i < y_j$ และคู่อันดับ (i, j) ไม่สอดคล้อง (concordant) กันเมื่อ (x_i, y_i) และ (x_j, y_j) ไม่เป็นไปในทางเดียวกันคือ $x_i > x_j, y_i < y_j$ หรือ $x_i < x_j, y_i > y_j$ โดยที่ x_i และ x_j คือผลลัพธ์ที่ได้จากการทำนายอันดับที่ i และ j และ y_i และ y_j คือผลลัพธ์ที่แท้จริงอันดับที่ i และ j

เมื่อทำการจับคู่อันดับครบทุกคู่ที่เป็นไปได้และแต่ละคู่อันดับมีการบ่งบอกค่าความสอดคล้องหรือค่าความไม่สอดคล้องครบทุกคู่อันดับแล้วจึงทำการนับผลรวมของจำนวนคู่อันดับที่มีค่าความสอดคล้องและจำนวนคู่อันดับที่ไม่สอดคล้องเพื่อนำมาพิจารณาค่าเคนดอลล์ตามสมการที่ 2.9

$$\tau = \frac{c-d}{\binom{n}{2}} \quad (2.9)$$

โดยที่
$$\binom{n}{2} = \frac{n(n-1)}{2} \quad (2.10)$$

- เมื่อ c คือผลรวมของจำนวนคู่ที่สอดคล้องกัน
 d คือผลรวมของจำนวนคู่ที่ไม่สอดคล้องกัน
 n คือจำนวนอันดับทั้งหมด

เมื่อได้ค่าเคนดอลล์แล้วจะนำค่าที่ได้ไปสู่การแปลความหมายของค่าเคนดอลล์กล่าวคือเมื่อค่าเคนดอลล์ที่ได้มีค่าเข้าใกล้หรือเท่ากับ 1 แสดงว่าอันดับที่ได้จากการทำนายนั้นให้ผลลัพธ์ที่สอดคล้องกับอันดับของผลลัพธ์ที่แท้จริง หากค่าเคนดอลล์ที่ได้มีค่าใกล้เคียงหรือเข้าใกล้ -1 แสดงว่าอันดับที่ได้จากการทำนายนั้นให้ผลลัพธ์ที่ไม่สอดคล้องกับผลลัพธ์ที่แท้จริง สามารถอธิบายได้ตามตัวอย่าง

ตัวอย่าง การคำนวณหาค่าความสอดคล้องของเคนดอลล์ พิจารณาจากเซตของคลาส $I = \{i_1, i_2, i_3, i_4\}$ ซึ่งมีการคำนวณอันดับที่แท้จริงกับอันดับที่ได้จากการทำนายดังตารางที่ 2.1

ตารางที่ 2.1 อันดับที่แท้จริงเทียบกับอันดับที่เกิดจากการทำนาย

	อันดับจากการทำนาย	อันดับจริง
i_1	1	1
i_2	2	4
i_3	3	2
i_4	4	3

จากตารางที่ 2.1 จำนวนผลลัพธ์มีทั้งหมด 4 คลาสที่แตกต่างกัน ดังนั้นจึงได้คู่ลำดับของผลลัพธ์ทั้งหมด 6 คู่ดังต่อไปนี้คือ (i_1, i_2) (i_1, i_3) (i_1, i_4) (i_2, i_3) (i_2, i_4) (i_3, i_4)

วิธีการหาค่าความสอดคล้องเช่นคู่อันดับ (i_1, i_2) เพราะฉะนั้นจะได้ค่า $i = 1$ และ $j = 2$ พิจารณาว่า $x_1 = 1$, $x_2 = 2$, $y_1 = 1$ และ $y_2 = 4$ เพราะฉะนั้น $x_1 < x_2$ และ $y_1 < y_2$ หมายความว่าคู่อันดับ (i_1, i_2) มีความสอดคล้องกัน เมื่อพิจารณาทุกคู่อันดับแล้วจะได้ค่าความสอดคล้องตามตารางที่ 2.2

ตารางที่ 2.2 ค่าความสอดคล้องกันของแต่ละคู่ลำดับของผลลัพธ์

	อันดับจากการทำนาย	อันดับจริง	ค่าความสอดคล้อง
(i_1, i_2)	$1 < 2$	$1 < 4$	สอดคล้อง
(i_1, i_3)	$1 < 3$	$1 < 2$	สอดคล้อง
(i_1, i_4)	$1 < 4$	$1 < 3$	สอดคล้อง
(i_2, i_3)	$2 < 3$	$4 > 2$	ไม่สอดคล้อง
(i_2, i_4)	$2 < 4$	$4 > 3$	ไม่สอดคล้อง
(i_3, i_4)	$3 < 4$	$2 < 3$	สอดคล้อง

เมื่อได้ค่าความสอดคล้องครบทุกคู่ลำดับของผลลัพธ์แล้วจึงนำมาหาค่าเคนดอลล์ โดยที่ค่า $c = 4$, $d = 2$ และ $n = 4$ นำไปแทนค่าในสมการที่ 2.9 จะได้ค่าเคนดอลล์เท่ากับ $\tau = \frac{4-2}{6} = 0.33$

2.6 งานวิจัยที่เกี่ยวข้อง

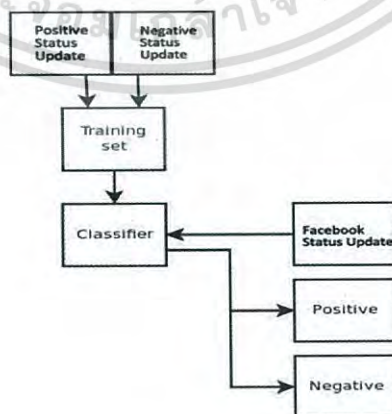
2.6.1 การวิเคราะห์ความรู้สึกของสถานะบนเฟสบุ๊คด้วยตัวจำแนกประเภทแบบเบย์[5]

การขยายตัวของข้อความแสดงความคิดเห็นที่เติบโตขึ้นซึ่งรวบรวมอยู่ในเว็บไซต์ที่อนุญาตให้ผู้ใช้เผยแพร่แบ่งปันประสบการณ์ได้ ด้วยขนาดของข้อมูลที่มีจำนวนมากทำให้เป็นที่สนใจเพื่อศึกษาการทำเหมืองข้อมูลและวิเคราะห์ความรู้สึกซึ่งได้รับความนิยมเป็นอย่างมากในปัจจุบัน เนื่องจากการทำการตลาดในปัจจุบันต้องการขยายฐานลูกค้าจากการวิเคราะห์ความพึงพอใจเกี่ยวกับสินค้าและบริการ

จากการศึกษาพบว่ามีกรวิเคราะห์ความรู้สึกของข้อความแสดงความคิดเห็นโดยแสดงอยู่บนเว็บไซต์ทวิตเตอร์ (www.twitter.com) แต่ด้วยข้อจำกัดของทวิตเตอร์ที่อนุญาตให้ผู้ใช้พิมพ์ข้อความได้เพียง 140 ตัวอักษร ทำให้ผู้ใช้งานทวิตเตอร์ใช้ตัวย่อและสัญลักษณ์มากขึ้น ด้วยเหตุนี้ทำให้การวิเคราะห์ความรู้สึกของข้อความแสดงความคิดเห็นในทวิตเตอร์เป็นเรื่องที่ยากด้วยความแตกต่างของเว็บไซต์เฟสบุ๊ค (www.facebook.com) ที่อนุญาตให้ผู้ใช้อัปเดต (Update) สถานะได้ถึง 5,000 ตัวอักษร ทำให้มีความชัดเจนในข้อความและด้วยแนวโน้มของจำนวนผู้ใช้งานเฟสบุ๊คที่เพิ่มขึ้นจึงถือเป็นตัวอย่างที่ดีสำหรับทำการวิเคราะห์ความรู้สึกของผู้ใช้ที่น่าสนใจ

การวิเคราะห์ความรู้สึกจัดเป็นการประมวลผลภาษาธรรมชาติ โดยเริ่มจากการจัดหมวดหมู่ข้อความเป็นข้อความที่มีทัศนคติที่ไม่เป็นกลางและข้อความที่มีทัศนคติเป็นกลางก่อน หลังจากนั้นจัดหมวดหมู่ของข้อความที่มีทัศนคติที่ไม่เป็นกลางเป็นทัศนคติเชิงบวกและทัศนคติเชิงลบ นอกจากนี้ยังทำการวิเคราะห์อิมโอดีคอนในข้อความแบ่งเป็น 2 ประเภท (1) ข้อความที่ใช้อิมโอดีคอน เช่น :-) :) =) หมายถึงความสุขหรือความรู้สึกเชิงบวก (2) ข้อความที่ใช้อิมโอดีคอน เช่น :-(:(=(;(หมายถึงไม่ชอบหรือความรู้สึกเชิงลบ

วิธีการหลักที่ใช้ในการวิเคราะห์ความรู้สึกคือวิธีการของตัวจำแนกประเภทแบบเบย์ (Naïve Bayes Classifier) สำหรับแบ่งสถานะเป็นทัศนคติเชิงบวกหรือทัศนคติเชิงลบ



รูปที่ 2.4 วิธีการหลักของตัวจำแนกประเภทแบบเบย์ (Naive Bayes Classifier)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวจำแนกประเภทแบบเบย์มีพื้นฐานจากกฎของเบย์ โดยความน่าจะเป็นแบบมีเงื่อนไขจะถูกปรับเปลี่ยนให้อยู่ในรูปแบบวิธีการที่สะดวกต่อการคำนวณ ซึ่งเงื่อนไขที่เป็นไปได้ที่จะเกิดเหตุการณ์ X ถ้าเกิดเหตุการณ์ Y ขึ้น อาจเขียนอยู่ในรูป $P(X|Y)$ มองในทางตรงข้ามถ้าเหตุการณ์ X ได้เกิดขึ้นแล้ว และสนใจจะหาความน่าจะเป็นที่จะเกิดเหตุการณ์ Y โดยทั้ง 2 เหตุการณ์เป็นอิสระต่อกัน สามารถใช้ทฤษฎีของเบย์ในการแก้ปัญหาดังกล่าวและอาจเขียนอยู่ในรูป $P(X|Y) = P(X)P(Y|X)/P(Y)$ โดยตัวจำแนกประเภทแบบเบย์จะทำการวิเคราะห์ความรู้สึกที่ถูกรวบรวมไว้ในคลังข้อมูล (Corpus)

คลังข้อมูล (Corpus) หมายถึงการเก็บรวบรวมงานเขียนหรือบันทึกที่หมายถึงเหตุเพื่อใช้ในการวิเคราะห์ทางภาษา โดยโปรแกรมที่ใช้ในการวิเคราะห์ความรู้สึกในงานเขียนที่ถูกรวบรวมจะถูกจำแนกเป็น 2 กลุ่มเท่าๆ กัน คือ ความรู้สึกเชิงบวกและความรู้สึกเชิงลบ จากการอัปเดตสถานะของผู้ใช้งานเฟซบุ๊ก ประมาณ 5,000 – 10,000 สถานะ

เมื่อนำความน่าจะเป็นแบบมีเงื่อนไขโดยใช้ทฤษฎีของเบย์ (Naive Bayes) มาสร้างตัวจำแนกประเภท จึงมีสูตรดังนี้

$$P(\text{sentiment} | \text{sentence}) = P(\text{sentiment}) P(\text{sentence} | \text{sentiment}) / P(\text{sentence})$$

นอกจากนี้ถ้าต้องการจัดอันดับความน่าจะเป็นของกลุ่มข้อมูลที่มีค่าความน่าจะเป็นเท่ากันสามารถใช้สมมติฐานที่เป็นอิสระต่อกันโดย $P(\text{sentence} | \text{sentiment})$ เป็นส่วนหนึ่งของ $P(\text{token} | \text{sentiment})$ โดยสามารถคำนวณ $P(\text{token} | \text{sentiment})$ ได้จากการบวกเพิ่ม 1 และการนับ token เรียกการกระทำนี้ว่า Add one หรือ Laplace smoothing

$$\text{count}(\text{this token in class}) + 1 / \text{count}(\text{all tokens in class}) + \text{count}(\text{all tokens})$$

ในขั้นตอนการวิเคราะห์ใช้ตัวจำแนก 3 ประเภท ได้แก่ เบย์ (Naive Bayes) ร็อคซิโอ (Rocchio) และเพอร์เซ็ปตรอน (Perceptron) เพื่อเปรียบเทียบประสิทธิภาพในการวิเคราะห์สถานะเฟซบุ๊กว่าเป็นทัศนคติเชิงบวกหรือเชิงลบ โดยรวบรวมสถานะ 7,000 สถานะจากผู้ใช้งาน 90 คน

โดยตัวจำแนกประเภทแต่ละประเภทจะใช้ตัวชี้วัด 3 ค่าในการเปรียบเทียบประสิทธิภาพ ได้แก่ ค่าความถูกต้อง (Precision) ค่าความครบถ้วน (Recall) และค่าประสิทธิภาพพื้นฐานในการจัดกลุ่ม (F-score)

$$\text{Precision} = \frac{tp}{tp+fp} \quad (2.11)$$

$$\text{Recall} = \frac{tp}{tp+fn} \quad (2.12)$$

$$F - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.13)$$

จะได้ตัวจำแนกประเภทแบบเบย์ (Naive Bayes Classifier) ที่มีค่าความถูกต้องเท่ากับ 0.77 ค่าความครบถ้วนเท่ากับ ตัวจำแนก 0.72 และค่าประสิทธิภาพพื้นฐานในการจัดกลุ่มเท่ากับ 0.68 ประเภทเบย์ร็อคซิโอ (Rocchio Classifier) ที่มีค่าความถูกต้องเท่ากับ 0.75 ค่าความครบถ้วนเท่ากับ และตัวจำแนกประเภทแบบเพอร์ 0.74 และค่าประสิทธิภาพพื้นฐานในการจัดกลุ่มเท่ากับ 0.73

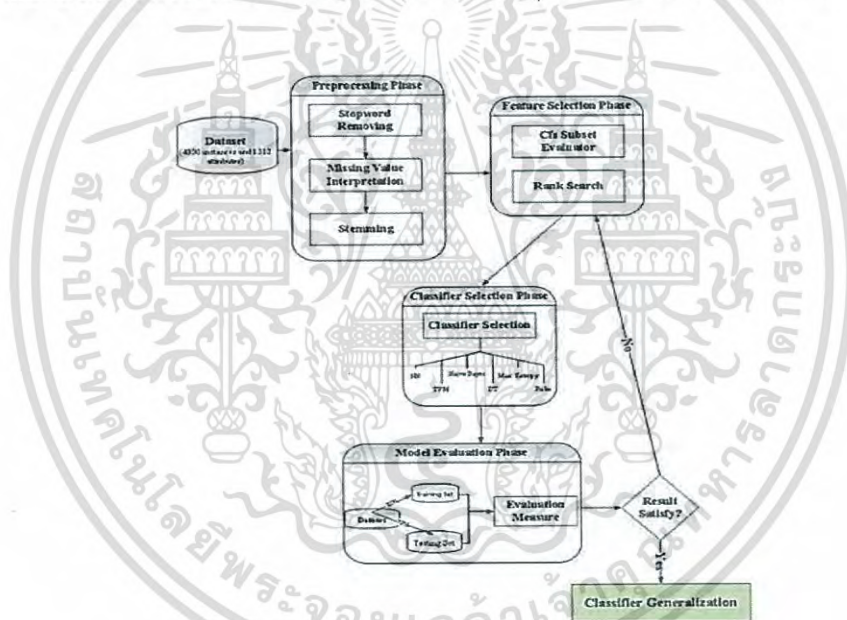
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เซ็ปตรอน(Perceptron Classifier) ที่มีค่าความถูกต้องเท่ากับ 0.65 ค่าความครบถ้วนเท่ากับ 0.56 0.60 และค่าประสิทธิภาพพื้นฐานในการจัดกลุ่มเท่ากับ

ถ้าพิจารณาที่ค่า F-score แสดงให้เห็นว่าตัวจำแนกประเภทแบบร็อคซิโอมีประสิทธิภาพมากที่สุดและ ตัวจำแนกประเภทแบบเพอร์เซ็ปตรอนมีประสิทธิภาพน้อยที่สุด โดยตัวจำแนกประเภทแบบเบย์มีประสิทธิภาพใกล้เคียงกับตัวจำแนกประเภทแบบร็อคซิโ

2.6.2 การจำแนกแบบเบย์เป็นการจำแนกประเภทเอกสารที่ดีที่สุด[6]

การจำแนกประเภทที่เหมาะสมของข้อมูลขนาดใหญ่ให้เป็นไปตามความต้องการ เป็นขั้นตอนสำคัญต่อความสำเร็จทางธุรกิจ อย่างไรก็ตามการจำแนกประเภทเป็นงานที่ใช้เวลานานและใช้ความละเอียดของมนุษย์ ซึ่งต้องจำแนกประเภทเอง การจำแนกประเภทแบบอัตโนมัติจึงมีการพัฒนามากขึ้น เช่น การจำแนกแบบเบย์ ซัพพอร์ตเวกเตอร์แมชชีน ต้นไม้ตัดสินใจ และเครือข่ายประสาทเทียม เป็นต้น ในวิธีการเหล่านี้การจำแนกแบบเบย์ได้รับการใช้อย่างแพร่หลาย นักวิจัยส่วนมากได้พิสูจน์ว่าการจำแนกแบบเบย์มีประสิทธิภาพเพียงพอที่จะจำแนกกลุ่มข้อความในหลายๆโดเมนได้



รูปที่ 2.5 ทฤษฎีที่เกี่ยวข้อง

การสร้างรูปแบบการจำแนกประเภทเอกสาร เริ่มต้นด้วย (1) การจัดเตรียมข้อมูล เป็นการสกัดข้อมูลที่ไม่เกี่ยวข้อง โดยกำหนด stop word ที่ไม่มีความสำคัญต่อเอกสาร หลังจากขั้นตอน stop word ข้อมูลที่หายไปจะถูกนำมาใช้ และมีการตีความค่าข้อมูล ค่าที่มีความหมายคล้ายกันแต่ไวยกรณ์ต่างกัน (bank กับ banks) ต้องนำมารวมให้เป็น 1 ลักษณะ ทำให้สามารถแสดงตัวอย่างของคำเหล่านั้นได้ดีกว่าและลดจำนวนชุดข้อมูลลง เวลาในการประมวลผลเร็วจึงขึ้น ซึ่งจะเป็นขั้นตอนที่ใช้เวลาในการดำเนินงานมากที่สุด (2) การคัดเลือกคุณลักษณะ เป็นสิ่งสำคัญมากของการทำเหมืองข้อมูล ช่วยลดคุณลักษณะที่ไม่จำเป็น โดยส่วนที่สำคัญจะถูกเก็บไว้และส่วนที่ไม่เกี่ยวข้องหรือมีการซ้ำจะถูกลบทิ้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพื่อจะประเมินประสิทธิภาพของขั้นตอนการจัดเตรียมข้อมูล ผลลัพธ์ที่ไม่จัดเตรียมข้อมูล และที่จัดเตรียมข้อมูลจะถูกเปรียบเทียบกัน ถ้าผลการทดสอบออกมาไม่ดี จะต้องมีการปรับค่าพารามิเตอร์ เช่น การเปลี่ยนเทคนิคที่ใช้เลือกคุณลักษณะ และสร้างรูปแบบใหม่อีกครั้ง โดยจะไม่หยุดทำขั้นตอนนี้นั้นจนกว่าจะได้การจำแนกประเภทที่ดี

การจำแนกในที่นี้แบ่งชุดข้อมูลออกเป็น 4 ประเภทคือ ธุรกิจ การเมือง กีฬา และการท่องเที่ยว เริ่มต้นด้วยการรับเอกสาร 1000 รายการสำหรับแต่ละประเภท รวมแล้วทั้งหมด 4000 รายการจะถูกแบ่งออกเป็นชุดข้อมูล 2 ชุดคือ ชุดการฝึกสอนและชุดทดสอบ โดย 30% (1200 เอกสาร) จะถูกนำไปอยู่ในชุดการฝึกสอน อีก 70% (2800 เอกสาร) จะอยู่ในชุดทดสอบ โดยจะไม่มีข้อมูลที่หายไป คุณลักษณะที่เป็นตัวเลขจะถูกนำไปวัดค่าน้ำหนักที่ใช้ประเมินค่าความสำคัญของค่านั้นๆ (TFIDF) และทำการทดสอบเปรียบเทียบประสิทธิภาพด้านความแม่นยำของเทคนิคการเลือกคุณลักษณะต่างๆ โดยใช้โปรแกรม Weka แต่น่าแปลกใจที่พบว่า ผลลัพธ์ของชุดข้อมูลที่มีการจัดเตรียมข้อมูล (95.5%) แย่กว่าที่ไม่ได้จัดเตรียมข้อมูล (96.9%) จึงจำเป็นต้องมีการปรับรูปแบบการจัดเตรียมข้อมูล เพื่อให้ได้รับผลลัพธ์ที่ดีกว่า โดยใช้การลดคุณลักษณะด้วยวิธี Correlation-based Feature Subset Selection (CFS) และการค้นหาแบบจัดลำดับ การใช้ Chi-square จึงถูกนำมาใช้ทำให้ค่าความถูกต้องเพิ่มขึ้น 0.1% ซึ่งช่วยให้การจำแนกดีขึ้น

เพื่อทดสอบว่าการจำแนกแบบเบย์ดีที่สุด จึงนำการจำแนกอีก 3 ประเภทคือ ซัพพอร์ตเวกเตอร์แมชชีน เครื่องข่ายประสาทเทียม และต้นไม้ตัดสินใจมาเปรียบเทียบ ผลลัพธ์ที่ได้คือ การจำแนกแบบเบย์ดีที่สุด แม้ว่า

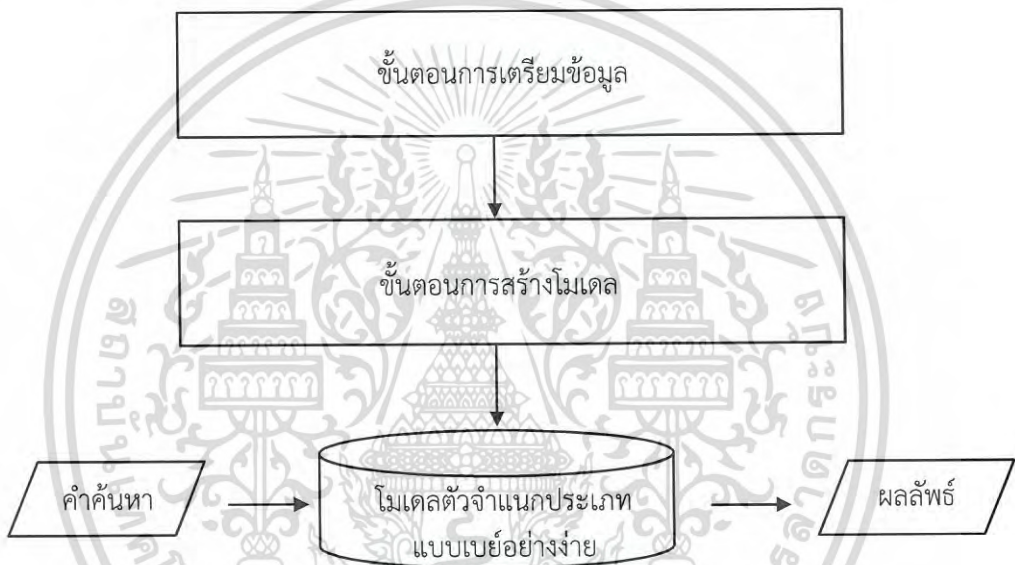
ซัพพอร์ตเวกเตอร์แมชชีนจะได้ผลลัพธ์เช่นเดียวกัน แต่เมื่อเทียบเวลาในการสร้างโมเดล ซัพพอร์ตเวกเตอร์แมชชีนใช้เวลา 2.69 วินาที ซึ่งการจำแนกแบบเบย์ใช้เพียง 0.19 วินาที

การจำแนกแบบเบย์ได้รับข้อกล่าวว่าเป็นตัวจำแนกประเภทเอกสารที่ดีที่สุด เมื่อเทียบร่วมกับตัวจำแนกอื่นๆ เนื่องจากสามารถเข้าใจได้ง่าย ระยะเวลาเหมาะสม มีความแม่นยำและการคำนวณที่มีประสิทธิภาพ หลังจากผ่านการดำเนินการของการเลือกคุณลักษณะที่แตกต่างกันและผ่านโปรแกรม Weka แสดงให้เห็นว่า การจัดเตรียมข้อมูลและการเลือกคุณลักษณะเป็น 2 ขั้นตอนที่สำคัญในการปรับปรุงคุณภาพของการทำเหมืองข้อมูล เนื่องจากผลการทำเหมืองข้อมูลได้รับผลกระทบโดยตรงจากคุณภาพของข้อมูล ขั้นตอนการจัดเตรียมข้อมูลจึงเป็นสิ่งสำคัญที่ทำให้ข้อมูลแม่นยำมากขึ้น และปรับปรุงเวลาที่ใช้ในการสร้างโมเดลให้ดีขึ้น สามารถสรุปได้ว่าลักษณะของข้อมูลแต่ละประเภทส่งผลต่อประสิทธิภาพการทำงานของแต่ละเทคนิค

บทที่ 3

โมเดลค้นหาโรงแรมและการออกแบบระบบ

เนื้อหาบทนี้ประกอบด้วยการสร้างโมเดลการค้นหาโรงแรมและการออกแบบระบบ ซึ่งขั้นตอนการสร้างโมเดลค้นหาโรงแรมประกอบด้วย 3 ขั้นตอนหลักคือ ขั้นตอนการเตรียมข้อมูล ขั้นตอนการสร้างโมเดลและขั้นตอนการใช้งานโมเดล ส่วนการออกแบบและพัฒนาระบบจะแสดงแผนภาพการทำงานของผู้ใช้งานระบบ (Use Case Diagram) ผังงาน (Flow Chart Diagram) แผนภาพแสดงลำดับการทำงานของระบบ (Sequence Diagram) การออกแบบฐานข้อมูล และตัวอย่างของส่วนติดต่อกับผู้ใช้ (Graphic User Interface)



รูปที่ 3.1 กระบวนการสร้างโมเดลการค้นหาโรงแรม

3.1 การสร้างโมเดลค้นหาโรงแรม

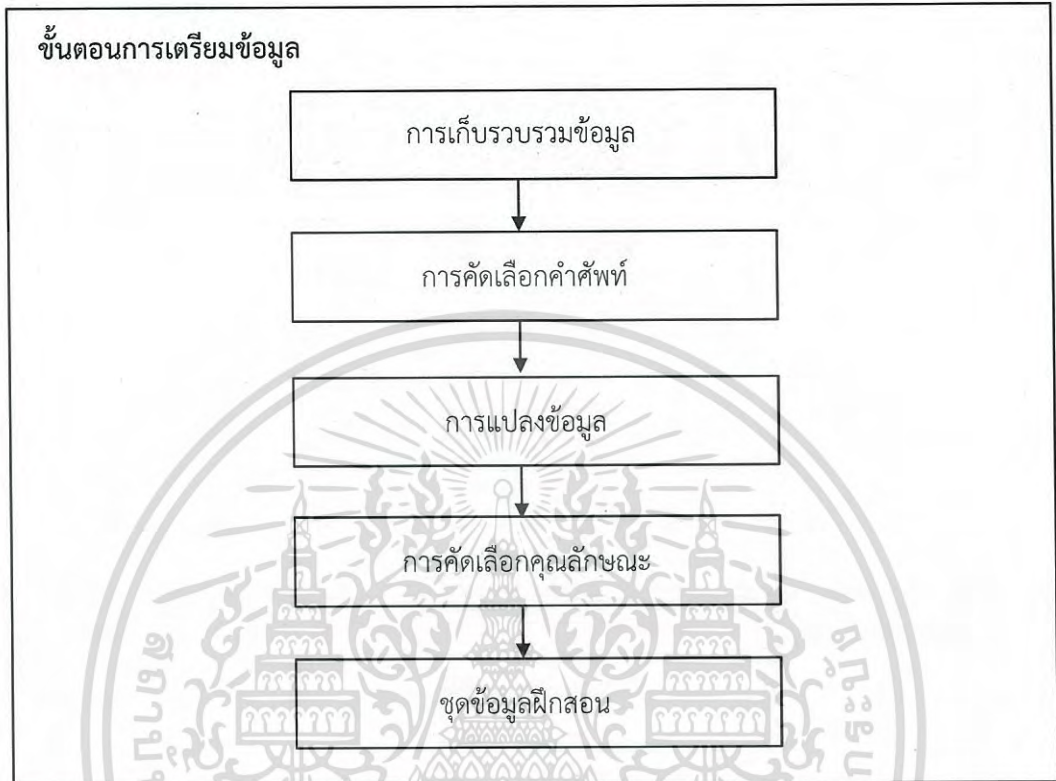
โดยภาพรวมของการสร้างโมเดลการค้นหาโรงแรมแสดงได้ดังรูปที่ 3.1

3.1.1 ขั้นตอนการเตรียมข้อมูล

ขั้นตอนสำหรับการเตรียมข้อมูล (Data preprocessing) เป็นขั้นตอนการเตรียมข้อมูลความคิดเห็นที่อยู่บนเว็บไซต์ ให้อยู่ในรูปแบบที่เหมาะสมเพื่อนำไปสร้างโมเดลในขั้นตอนถัดไป โดยขั้นตอนการเตรียมข้อมูลนี้สามารถแบ่งออกได้ 4 ขั้นตอนได้แก่ ขั้นตอนการเก็บรวบรวมข้อมูล (Data collection) ขั้นตอนการคัดเลือกคำศัพท์ (Feature extraction) ขั้นตอนการแปลงข้อมูล (Data transformation)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และขั้นตอนการคัดเลือกคุณลักษณะ (Feature selection) โดยขั้นตอนการเตรียมข้อมูล แสดงได้ดังรูปที่ 3.2



รูปที่ 3.2 ขั้นตอนการเตรียมข้อมูล

• ขั้นตอนการเก็บรวบรวมข้อมูล

ในขั้นตอนนี้เป็นการเก็บรวบรวมข้อมูลความคิดเห็นต่อโรงแรมต่างๆ จากเว็บไซต์ www.tripadvisor.com โดยคัดเลือกเฉพาะโรงแรมตัวอย่างในเขตการปกครองพิเศษเมืองพัทยา จังหวัดชลบุรีจำนวน 60 แห่งดังนี้

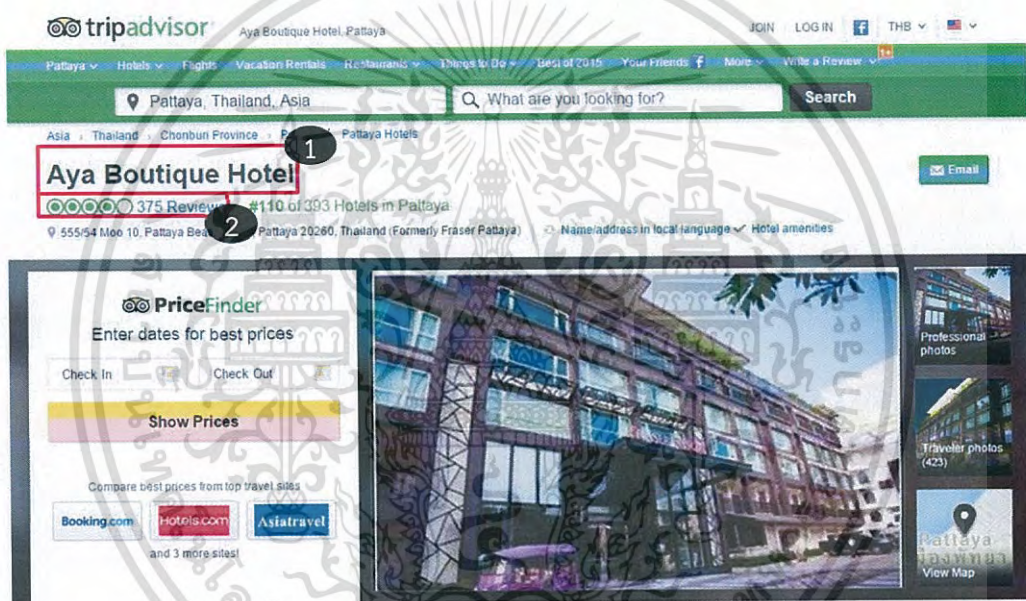
1. Baboona BeachFront Living
2. Basaya Beach Hotel & Resort
3. Bella Express
4. Citrus Parc Pattaya Hotel
5. D Varee Jomtien Beach
6. Eastiny Seven
7. Fairtex Sports Club Hotel
8. Grand Sole Pattaya Beach Hotel
9. Green Park Resort
10. KTK Royal Residence
11. Le Viman Resort

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลง บิดเบือน และดัดแปลงอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

12. Mercure Pattaya Hotel
13. Nova Platinum Hotel Pattaya
14. Pattaya Bay Resort
15. Red Planet Pattaya
16. Sabai Lodge
17. Sunshine Garden Resort
18. Sunshine Hotel & Residences
19. The Nova Gold Hotel Pattaya
20. Zing Resort & Spa
21. Amari Garden Pattaya
22. Amari Nova Suites Pattaya
23. A-One The Royal Cruise Hotel
24. Areca Lodge
25. Avalon Beach Resort
26. Aya Boutique Hotel
27. Bella Villa Prima Hotel
28. Centara Grand Modus Resort Pattaya
29. Courtyard South Pattaya
30. Grand Jomtien Palace
31. Intimate Hotel
32. LK The Empress
33. Mercure Pattaya Ocean Resort
34. Pattaya Discovery Beach Hotel
35. Pinnacle Grand Jomtien Resort
36. Prima Villa Hotel
37. Sea me Spring Hotel
38. Siam Bayshore
39. The Bayview Hotel
40. Woodlands Hotel & Resort
41. Amari Ocean Pattaya
42. Aya Boutique Hotel
43. Cape Dara Resort
44. Centara Grand Mirage Beach Resort
45. Centara Grand Phratamnak Pattaya
46. Dusit Thani Pattaya
47. Hard Rock Hotel Pattaya
48. Hilton Pattaya
49. Holiday Inn Pattaya

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

50. Hotel Baraquada Pattaya - MGallery Collection
51. OZO Pattaya
52. Pattaya Marriott Resort & Spa
53. Pullman Pattaya Hotel G
54. Rabbit Resort
55. Royal Cliff Grand Hotel
56. Royal Wing Suites & Spa
57. Sheraton Pattaya Resort
58. Siam@Siam Design Hotel Pattaya
59. Sugar Hut Resort
60. Wave Hotel



รูปที่ 3.3 ตัวอย่างข้อมูลโรงแรม

จากรูปที่ 3.3 คือตัวอย่างหน้าเว็บไซต์ www.tripadvisor.com ที่รวบรวมข้อมูลของโรงแรมตัวอย่าง โดยส่วนที่ 1 คือชื่อโรงแรมและส่วนที่ 2 คือระดับดาวของโรงแรม

สำหรับการคัดเลือกความคิดเห็นของผู้ใช้ที่มีต่อโรงแรมจะคัดเลือกข้อมูลความคิดเห็นที่มีระดับดาวตั้งแต่ 3 ถึง 5 ดาว ซึ่งบ่งบอกถึงคะแนนเฉลี่ยในทุกๆ ด้าน โดยระดับของดาวสามารถอธิบายได้ดังนี้คือ

- 5 ดาว คือ ยอดเยี่ยม (Excellent)
- 4 ดาว คือ ดีมาก (Good)
- 3 ดาว คือ ปานกลาง (Average)
- 2 ดาว คือ แย่ (Poor)
- 1 ดาว คือ แย่ที่สุด (Terrible)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับข้อมูลความคิดเห็นที่ใช้ในการสร้างโมเดลประกอบด้วยข้อความ 1,260 ความคิดเห็น และข้อความความคิดเห็นสำหรับทดสอบโมเดลจำนวน 540 ความคิดเห็น สำหรับตัวอย่างความคิดเห็นของผู้ใช้ที่มีต่อโรงแรมได้ดังรูปที่ 3.4 ในส่วนที่ 1 คือความคิดเห็นของผู้ใช้บริการที่มีต่อโรงแรมและส่วนที่ 2 คือดาวสำหรับความพึงพอใจที่มีต่อโรงแรมในแต่ละด้านซึ่งประกอบด้วย ด้านราคา (Value) ด้านห้องพัก (Rooms) ด้านสถานที่ (Location) ด้านความสะอาด (Cleanliness) ด้านคุณภาพการนอน (Sleep Quality) และด้านการบริการ (Service)

รูปที่ 3.4 ความคิดเห็นของผู้ใช้บริการโรงแรม

● ขั้นตอนการคัดเลือกคำศัพท์

ในขั้นตอนการคัดเลือกคำศัพท์ (Feature extraction) จะใช้วิธีการคัดเลือกคำศัพท์ที่ปรากฏอยู่ในแต่ละความคิดเห็น โดยเก็บคำศัพท์ที่ปรากฏในความคิดเห็นและทำให้อยู่ในรูปของรากศัพท์โดยการตัด -s, -es, -ly, -ing, -ed และตัดคำสรรพนาม คำบุพบทต่างๆ เช่น he, she, it, is, am, are, a, an, at, but, in, on, the เป็นต้น ซึ่งคำศัพท์ที่ได้จากข้อมูลความคิดเห็นมีทั้งหมด 8,835 คำ หลังจากนั้นจะนำคำศัพท์ที่ได้มาสร้างพจนานุกรมคำศัพท์ (Data dictionary) เพื่อใช้ในขั้นตอนต่อไป

ตารางที่ 3.1 ตัวอย่างการสร้างรากศัพท์ของคำ

คำศัพท์เดิม	รากศัพท์	ตัวอักษรที่ตัดออก
guests	guest	-s
goes	go	-es
perfectly	perfect	-ly
amazing	amaze	-ing
used	use	-ed

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- **ขั้นตอนการแปลงข้อมูล**

ในขั้นตอนการแปลงข้อมูล (Data transformation) จะทำการแปลงข้อมูลความคิดเห็นจากผู้ใช้ให้อยู่ในรูปแบบความถี่ของคำศัพท์ ซึ่งในการแปลงข้อมูลนี้จะต้องทำคำศัพท์จากข้อมูลความคิดเห็นให้อยู่ในรูปแบบของรากศัพท์เช่นกัน หลังจากนั้นจึงนำมาเปรียบเทียบกับพจนานุกรมคำศัพท์ที่ได้จากขั้นตอนการคัดเลือกคำศัพท์โดยจะนับความถี่ของคำศัพท์ที่ปรากฏอยู่ในความคิดเห็น ดังตารางที่ 3.2

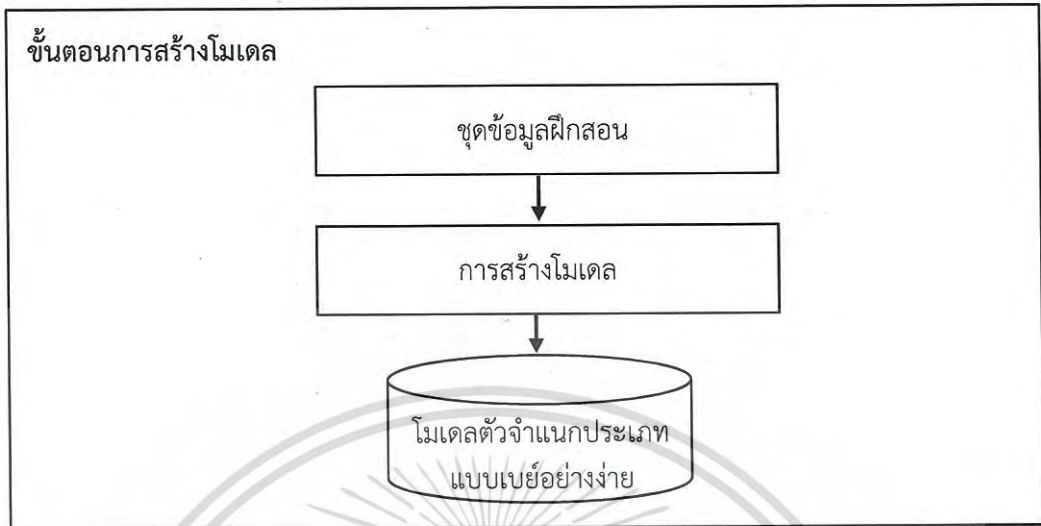
ตารางที่ 3.2 ตัวอย่างการแปลงข้อมูลความคิดเห็น

โรงแรมที่	ความถี่ของการปรากฏคำศัพท์					
	abdominal	ability	able	about	...	zone
1	0	7	1	12	...	2
2	5	4	6	18	...	8
3	2	3	8	22	...	0
.
.
.
60	1	0	12	25	...	3

- **ขั้นตอนการคัดเลือกคุณลักษณะ**

จากขั้นตอนที่ผ่านมาจะได้ความถี่ของคำศัพท์แต่ละคำที่ปรากฏในความคิดเห็นของแต่ละโรงแรมซึ่งมีคำศัพท์จำนวนมากส่งผลให้ใช้เวลามากในการสร้างโมเดล ดังนั้นจึงต้องทำการคัดเลือกคุณลักษณะ (Feature selection) จะทำการลดจำนวนคำศัพท์เพื่อลดเวลาสำหรับการสร้างโมเดล การค้นหาโดยใช้การคัดเลือกคุณลักษณะด้วยวิธี Information Gain และทำการคัดเลือกคำศัพท์ที่มีความสำคัญจำนวน 20% จากจำนวนคำศัพท์ทั้งหมด

3.1.2 ขั้นตอนการสร้างโมเดล



รูปที่ 3.5 ขั้นตอนการสร้างโมเดล

หลังจากขั้นตอนการคัดเลือกคุณลักษณะด้วยเทคนิค Information Gain จะได้ชุดข้อมูลฝึกสอนเพื่อนำมาสร้างโมเดลด้วยเทคนิคตัวจำแนกประเภทแบบเบย์อย่างง่าย ดังรูปที่ 3.5 ตารางที่ 3.3 แสดงค่าความน่าจะเป็นในการปรากฏคำศัพท์ของแต่ละโรงแรมที่สร้างจากโมเดลตัวจำแนกประเภทแบบเบย์อย่างง่าย

ตารางที่ 3.3 ค่าความน่าจะเป็นในการปรากฏคำศัพท์ของแต่ละโรงแรม

โรงแรมที่	คำศัพท์					
	access	always	apartment	appreciate	...	wifi
1	0.2516	0.2132	0.1764	0.2154	...	0.2313
2	0.2914	0.2718	0.2352	0.2671	...	0.2496
3	0.2362	0.2134	0.1677	0.2537	...	0.3572
.
.
.
60	0.2114	0.2038	0.2291	0.2154	...	0.3215

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.3 ขั้นตอนการใช้งานโมเดล

ในขั้นตอนการนำโมเดลไปใช้ จะนำค่าความน่าจะเป็นที่ได้ในขั้นตอนที่ผ่านมาใช้ในการคำนวณ เมื่อทำการป้อนตัวอย่างชุดคำค้นหาระบบจะทำการประมวลผลการค้นหาโดยใช้ค่าความน่าจะเป็นจากคำศัพท์ที่ตรงกับชุดคำค้นหามาคำนวณ ซึ่งการคำนวณค่าความน่าจะเป็นของแต่ละโรงแรมที่ปรากฏชุดของคำค้นหาในความคิดเห็นของโรงแรมนั้นๆ คำนวณได้จากสมการที่ 3.1

$$P(Hotel_k|keywords) = P(Hotel_k) \prod_{i=1}^n P(keyword_i|Hotel_k) \quad (3.1)$$

โดยที่ $P(Hotel_k|keywords)$ คือความน่าจะเป็นของโรงแรมที่ k เมื่อปรากฏคำค้นหาในความคิดเห็นที่มีต่อโรงแรม k

$P(Hotel_k)$ คือค่าความน่าจะเป็นของความคิดเห็นที่มีต่อโรงแรม k

$P(keyword_i|Hotel_k)$ คือความน่าจะเป็นของคำค้นหาที่ i ที่ปรากฏอยู่ในความคิดเห็นที่มีต่อโรงแรม k



รูปที่ 3.6 ขั้นตอนการใช้งานโมเดล

เมื่อคำนวณค่าความน่าจะเป็นได้ครบทุกโรงแรมแล้วจึงนำผลลัพธ์ที่ได้มาเรียงลำดับรายชื่อโรงแรมด้วยความน่าจะเป็นจากมากไปน้อย โดยจะยกตัวอย่างความน่าจะเป็นด้านความสะอาดจะแสดงดังตารางที่ 3.4

ตารางที่ 3.4 ตัวอย่างความน่าจะเป็นด้านความสะอาด

โรงแรม	ความน่าจะเป็น
Basaya Beach Hotel & Resort	0.0053714541
Centara Grand Modus Resort Pattaya	0.0051272971
Mercure Pattaya Ocean Resort	0.0088867356
Royal Cliff Grand Hotel	0.0179960520
Sheraton Pattaya Resort	0.0115374219

จากตารางที่ 3.4 จะได้ความน่าจะเป็นของแต่ละโรงแรมซึ่งผลลัพธ์ที่ได้ระบบจะแสดงลำดับโรงแรมตามความน่าจะเป็นโดยทำการเรียงลำดับจากมากไปน้อยดังตารางที่ 3.5

ตารางที่ 3.5 ตัวอย่างผลลัพธ์ที่ได้จากการเรียงลำดับ

ลำดับ	โรงแรม
1	Royal Cliff Grand Hotel
2	Sheraton Pattaya Resort
3	Mercure Pattaya Ocean Resort
4	Basaya Beach Hotel & Resort
5	Centara Grand Modus Resort Pattaya

3.2 การวิเคราะห์และออกแบบระบบ

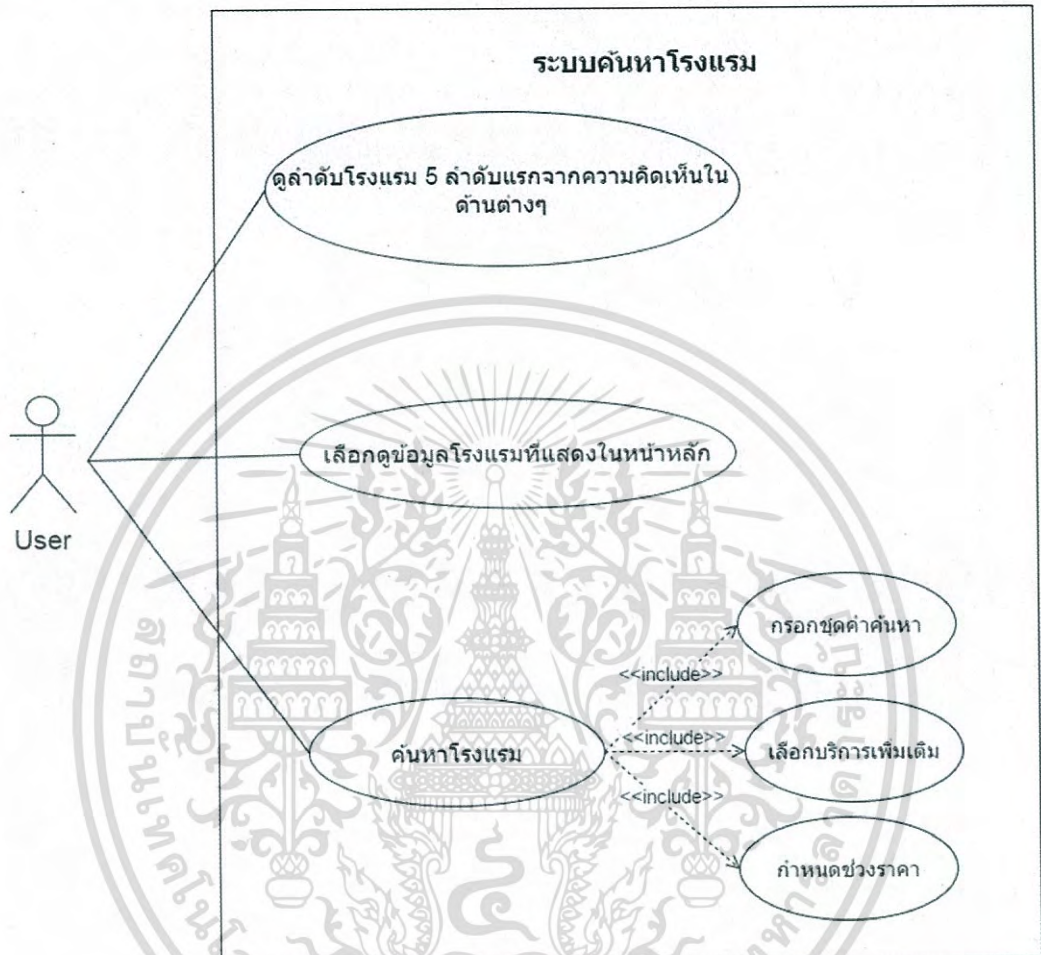
จากการวิเคราะห์ระบบอัจฉริยะในการค้นหาโรงแรม สามารถนำมาเขียนแผนภาพการทำงานของผู้ใช้งาน (Use Case Diagram) ผังงาน (Flow Chart Diagram) แผนภาพแสดงลำดับการทำงานของระบบ (Sequence Diagram) การออกแบบฐานข้อมูล และตัวอย่างของส่วนติดต่อกับผู้ใช้ (Graphic User Interface) เพื่อแสดงความสามารถและขั้นตอนการทำงานของระบบ

3.2.1 แผนภาพการทำงานของผู้ใช้งาน (Use Case Diagram)

โดยผู้ใช้งานระบบสามารถใช้งานระบบได้ดังรูปที่ 3.7

- ดูลำดับโรงแรม 5 ลำดับแรกจากความคิดเห็นในด้านต่างๆ
ผู้ใช้งานระบบสามารถดูลำดับโรงแรม 5 ลำดับที่ดีที่สุดในด้านต่างๆ โดยจะเรียงจากคะแนนที่ผู้เข้าพักจริงได้ให้ไว้ต่อโรงแรมจากมากไปน้อย และสามารถคลิกเข้าไปดูรายละเอียดของโรงแรมเพิ่มเติมได้
- เลือกดูข้อมูลโรงแรม
ผู้ใช้งานระบบสามารถเลือกดูข้อมูล รายละเอียดและรูปภาพของโรงแรมที่แสดงในหน้าแรกของเว็บแอปพลิเคชัน
- กรอกชุดคำค้นหา
ผู้ใช้งานระบบสามารถกรอกชุดคำค้นหาที่ต้องการ ระบบจะทำการประมวลผลการค้นหา โดยค้นหาจากชุดคำค้นหาที่ปรากฏในความคิดเห็นที่มีต่อโรงแรมด้วยโมเดลตัวจำแนกประเภทแบบเบย์อย่างง่าย
- เลือกบริการเพิ่มเติม
ผู้ใช้งานระบบสามารถเลือกบริการของโรงแรมได้ว่าต้องการโรงแรมที่มีสิ่งอำนวยความสะดวกหรือบริการเพิ่มเติม ตัวอย่างเช่น ต้องการโรงแรมที่มีสระว่ายน้ำ ฟิตเนส สปา เป็นต้น
- กำหนดช่วงราคา
ผู้ใช้งานระบบสามารถเลือกช่วงราคาของโรงแรมที่ต้องการได้ว่าต้องการค้นหาโรงแรมที่มีราคาอยู่ในช่วงที่กำหนด โดยระบบมีให้เลือกช่วงราคา 3 ช่วงราคา คือ 500 ถึง 1,500 บาท 1,501 ถึง 2,500 บาท และ 2,500 บาทขึ้นไป
หลังจากผู้ใช้งานระบบทำการค้นหาแล้ว ผู้ใช้ระบบสามารถดูลำดับโรงแรมที่ได้จากการค้นหา

โดยจะเรียงจากค่าความน่าจะเป็นของชุดคำค้นหาที่ปรากฏในความคิดเห็นที่มีต่อโรงแรม จากมากไปน้อย และสามารถคลิกเข้าไปดูรายละเอียดของโรงแรมเพิ่มเติมได้ เช่น ที่อยู่ เบอร์โทรศัพท์ ติดต่อ ราคาและบริการ เป็นต้น

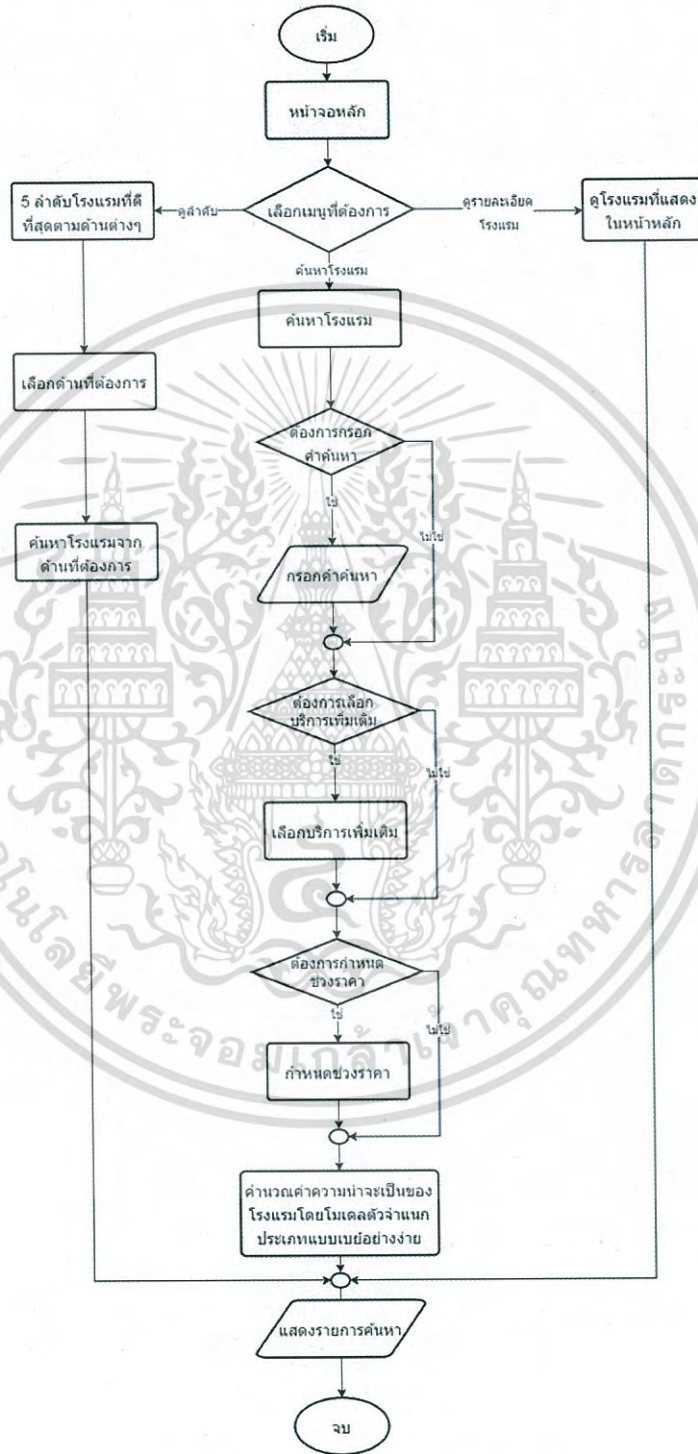


รูปที่ 3.7 แผนภาพการทำงานของผู้ใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2 ผังงาน (Flow Chart Diagram)

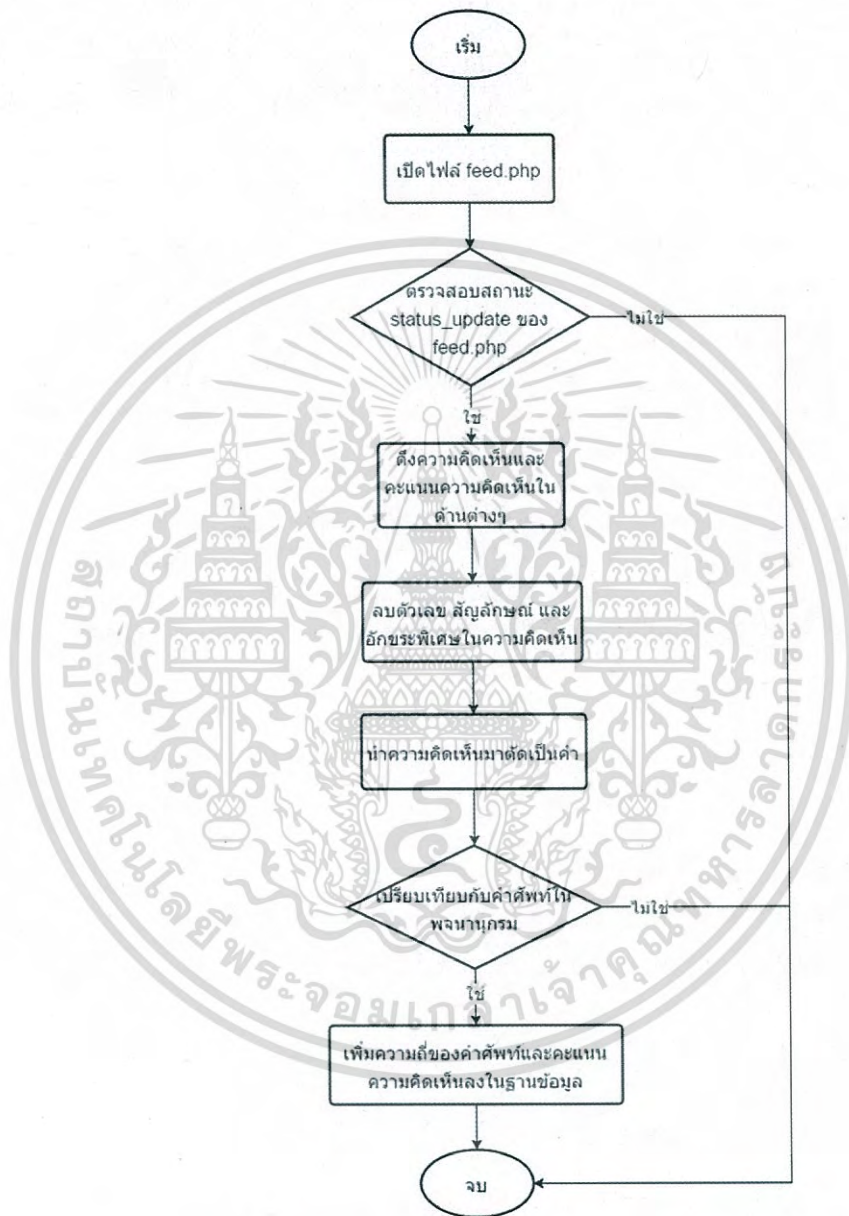
ผังงานในระบบแบ่งออกเป็น 2 ส่วน คือส่วนที่แสดงถึงลำดับขั้นตอนในการทำงานของระบบ ในส่วนของผู้ใช้บนหน้าเว็บเพจตั้งแต่รับข้อมูล คำนวณจนถึงแสดงผลลัพธ์ แสดงได้ดังรูปที่ 3.8



รูปที่ 3.8 ผังงานการทำงานของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

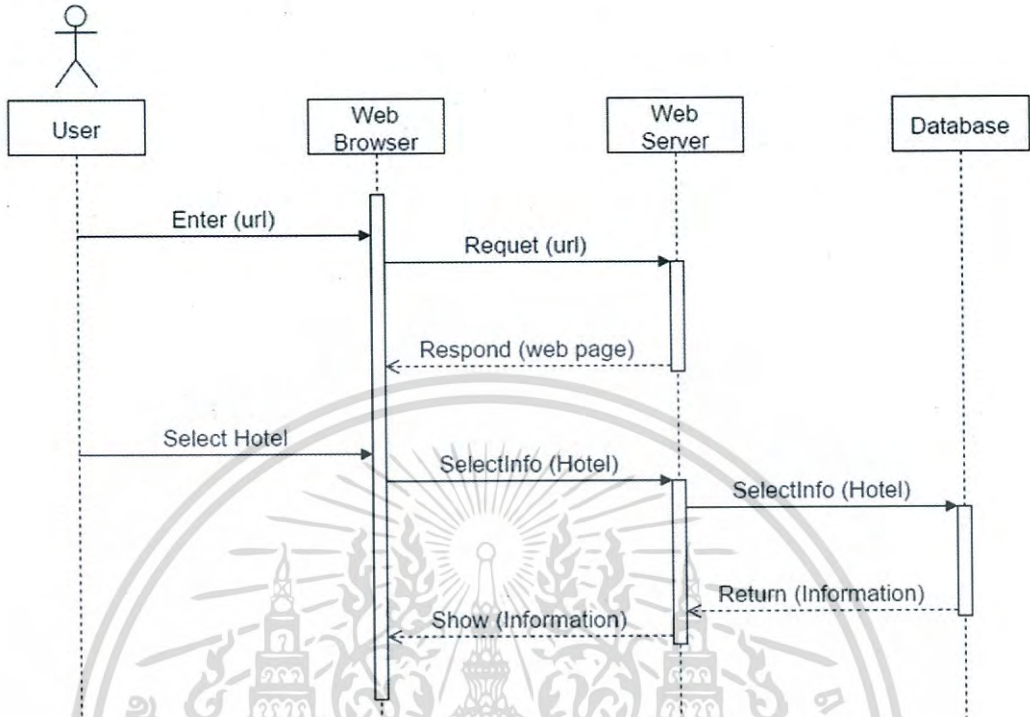
และส่วนที่แสดงการทำงานของเอเจนต์ ซึ่งทำงานอัตโนมัติในฝั่งเครื่องเซิร์ฟเวอร์ โดยจะทำการดึงข้อมูลความคิดเห็นมาทำการตัดคำและเพิ่มความถี่ให้กับคำศัพท์ในพจนานุกรมในกรณีที่คำศัพท์เหมือนกัน แสดงได้ดังรูปที่ 3.9



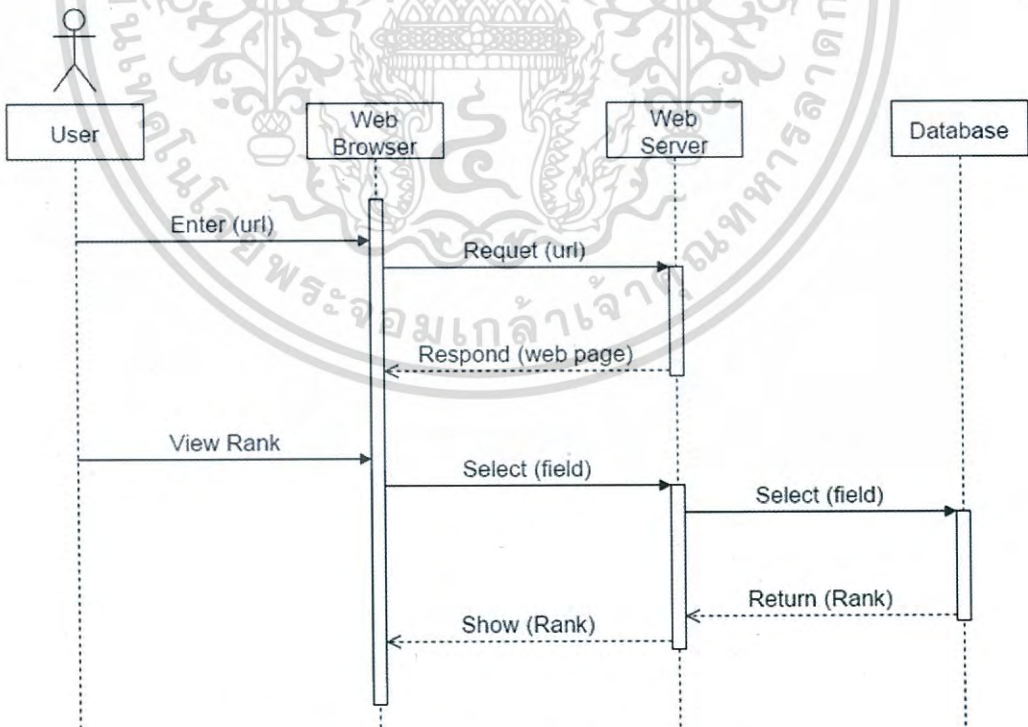
รูปที่ 3.9 ฝั่งงานแสดงการทำงานของเอเจนต์ที่ฝั่งเซิร์ฟเวอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.3 แผนภาพแสดงลำดับการทำงานของระบบ (Sequence Diagram)

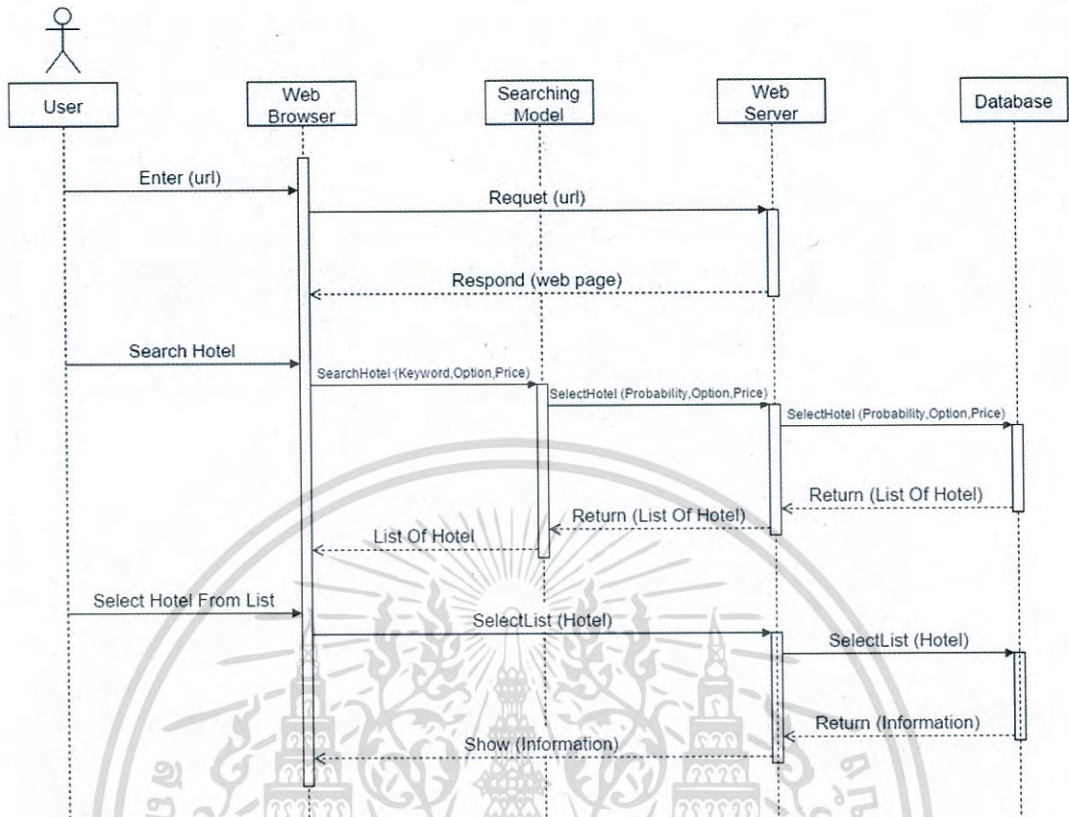


รูปที่ 3.10 แผนภาพแสดงการดูรายละเอียดโรงแรม



รูปที่ 3.11 แผนภาพแสดงการดูลำดับโรงแรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

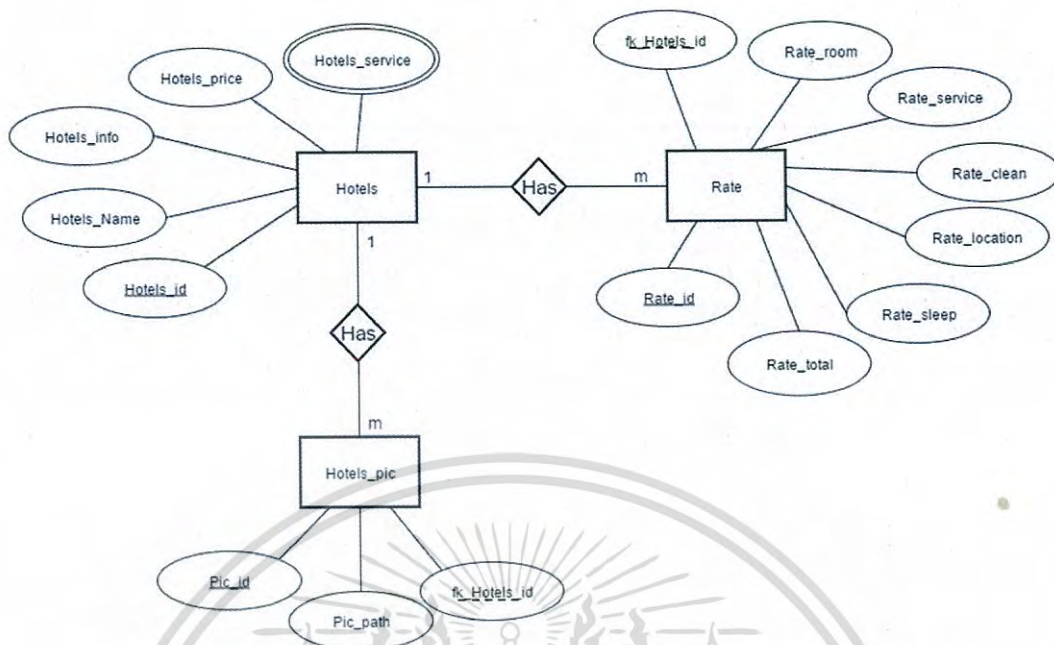


รูปที่ 3.12 แผนภาพแสดงการค้นหาโรงแรม

3.2.4 การออกแบบฐานข้อมูล

ระบบอัจฉริยะในการค้นหาโรงแรมมีการเก็บข้อมูลลงฐานข้อมูล ทำให้ต้องมีการออกแบบฐานข้อมูลเพื่อแสดงรูปแบบการเก็บข้อมูลในฐานข้อมูลและมองเห็นโครงสร้างงานชัดเจนยิ่งขึ้น

แบบจำลองความสัมพันธ์เอนทิตีเป็นแผนภาพที่ใช้ออกแบบจำลองข้อมูลที่ประกอบไปด้วยเอนทิตีและความสัมพันธ์ระหว่างข้อมูล (Relationship) ที่เกิดขึ้นทั้งหมดภายในระบบอัจฉริยะค้นหาโรงแรม ดังรูปที่ 3.13



รูปที่ 3.13 แบบจำลองความสัมพันธ์เอนทิตีของระบบอัจฉริยะในการค้นหาโรงแรม

ตารางที่ 3.6 ถึง 3.8 อธิบายถึงรายละเอียดตารางข้อมูลต่างๆ ของฐานข้อมูลที่อยู่ในระบบอัจฉริยะค้นหาโรงแรมซึ่งประกอบด้วย ชื่อฟิลด์ ประเภทข้อมูล รายละเอียดข้อมูลและในระบบอัจฉริยะในการค้นหาโรงแรม โดยตารางที่ 3.6 ใช้สำหรับเก็บข้อมูลของโรงแรม ได้แก่ ชื่อ รายละเอียด ราคา และบริการของโรงแรม ตารางที่ 3.7 ใช้สำหรับเก็บคะแนนความพึงพอใจในแต่ละด้านที่ได้จากความคิดเห็นของผู้เข้าพัก และตารางที่ 3.8 ใช้สำหรับเก็บที่อยู่ของไฟล์รูปภาพของแต่ละโรงแรม

ตารางที่ 3.6 รายละเอียดของตาราง Hotels

ชื่อฟิลด์	ประเภทข้อมูล	รายละเอียด
Hotels_id	NUMBER(4)	เป็นคีย์หลักที่ใช้เก็บรหัสโรงแรม
Hotels_name	VARCHAR(30)	เก็บชื่อโรงแรม
Hotels_info	VARCHAR(30)	เก็บรายละเอียดเพิ่มเติมของโรงแรม
Hotels_price	NUMBER(4)	เก็บราคาโรงแรม
Hotels_service	BINARY	เก็บบริการของโรงแรม

ตารางที่ 3.7 รายละเอียดของตาราง Rate

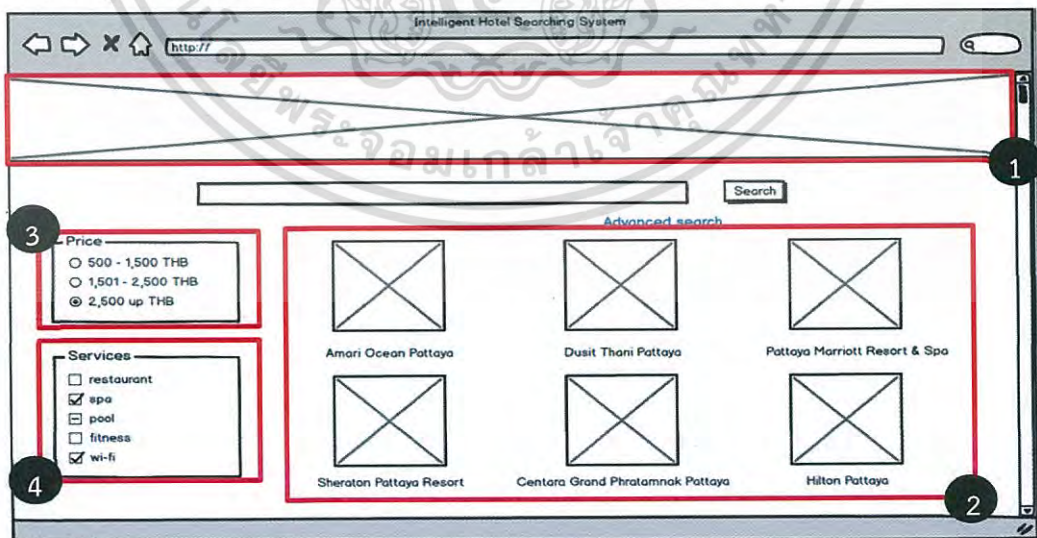
ชื่อฟิลด์	ประเภทข้อมูล	รายละเอียด
Rate_id	NUMBER(4)	เป็นคีย์หลักที่ใช้เก็บรหัสคะแนนความพึงพอใจ
Rate_room	NUMBER(2)	เก็บคะแนนความพึงพอใจด้านห้องพัก
Rate_service	NUMBER(2)	เก็บคะแนนความพึงพอใจด้านการบริการ
Rate_clean	NUMBER(2)	เก็บคะแนนความพึงพอใจด้านความสะอาด
Rate_location	NUMBER(2)	เก็บคะแนนความพึงพอใจด้านสถานที่
Rate_sleep	NUMBER(2)	เก็บคะแนนความพึงพอใจด้านคุณภาพการนอน
Rate_total	NUMBER(2)	เก็บคะแนนความพึงพอใจโดยรวม
fk_Hotels_id	NUMBER(4)	เป็นคีย์อ้างอิงที่อ้างถึงตาราง Hotels

ตารางที่ 3.8 รายละเอียดของตาราง Hotels_pic

ชื่อฟิลด์	ประเภทข้อมูล	รายละเอียด
Pic_id	NUMBER(4)	เป็นคีย์หลักที่ใช้เก็บรหัสของรูปภาพ
Pic_path	VARCHAR(2083)	เก็บชื่อของที่อยู่ของรูปภาพ
fk_Hotels_id	NUMBER(4)	เป็นคีย์อ้างอิงที่อ้างถึงตาราง Hotels

3.2.5 ตัวอย่างของส่วนติดต่อกับผู้ใช้ (Graphic User Interface)

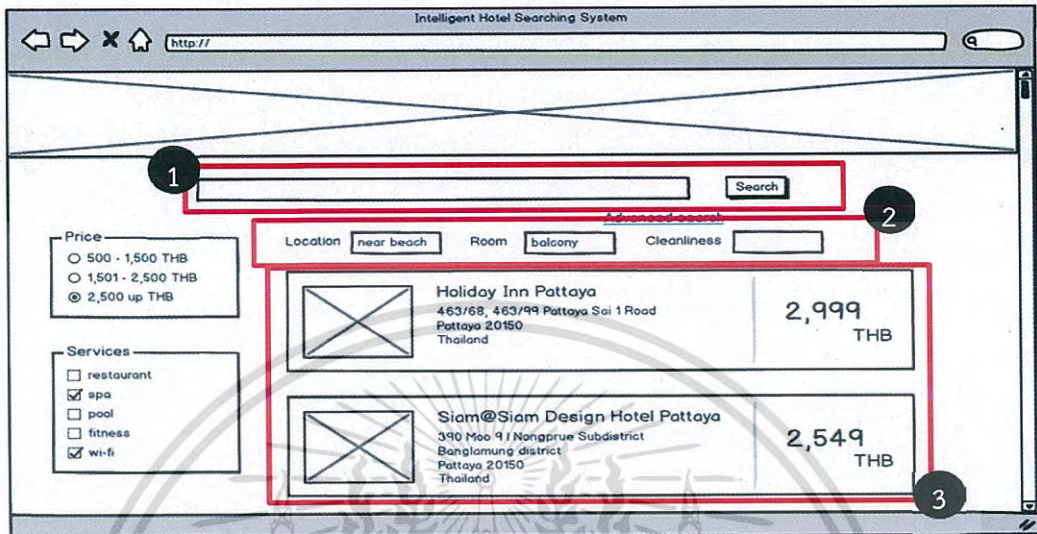
หน้าจอหลักประกอบไปด้วย ส่วนที่ 1 คือชื่อเว็บแอปพลิเคชัน ส่วนที่ 2 คือชื่อและรูปภาพโรงแรมที่แสดงในหน้าหลัก ส่วนที่ 3 คือช่วงราคาที่ใช้สามารถกำหนดได้ และส่วนที่ 4 คือบริการเพิ่มเติมที่ผู้ใช้สามารถเลือกได้ แสดงได้ดังรูปที่ 3.14



รูปที่ 3.14 หน้าหลักของระบบ

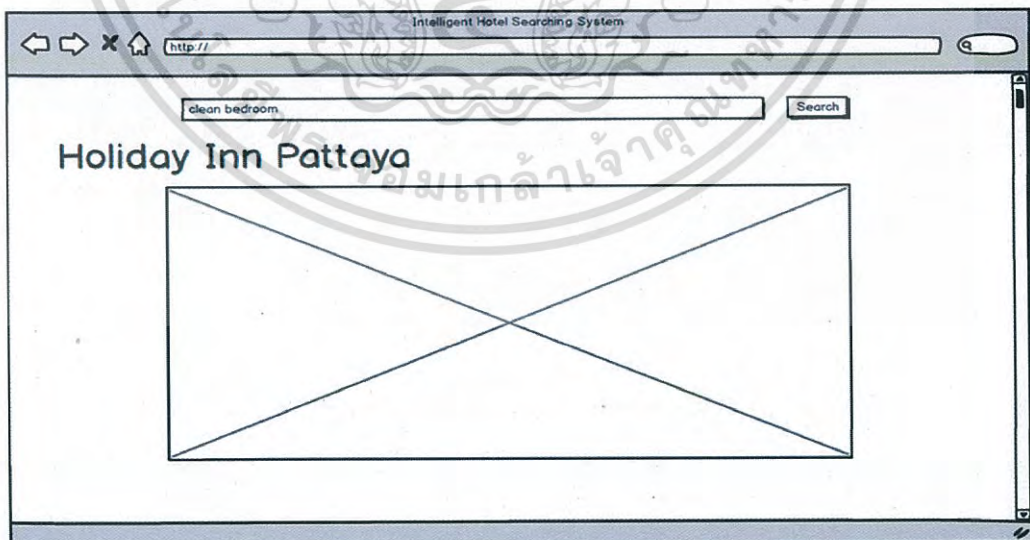
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หน้าจอกำหนดค้นหาโรงแรมประกอบด้วย ส่วนที่ 1 คือการค้นหาเบื้องต้น ส่วนที่ 2 คือการค้นหาขั้นสูง และส่วนที่ 3 คือชื่อ รูปภาพและข้อมูลโรงแรมที่ค้นหาแสดงได้ดังรูปที่ 3.15



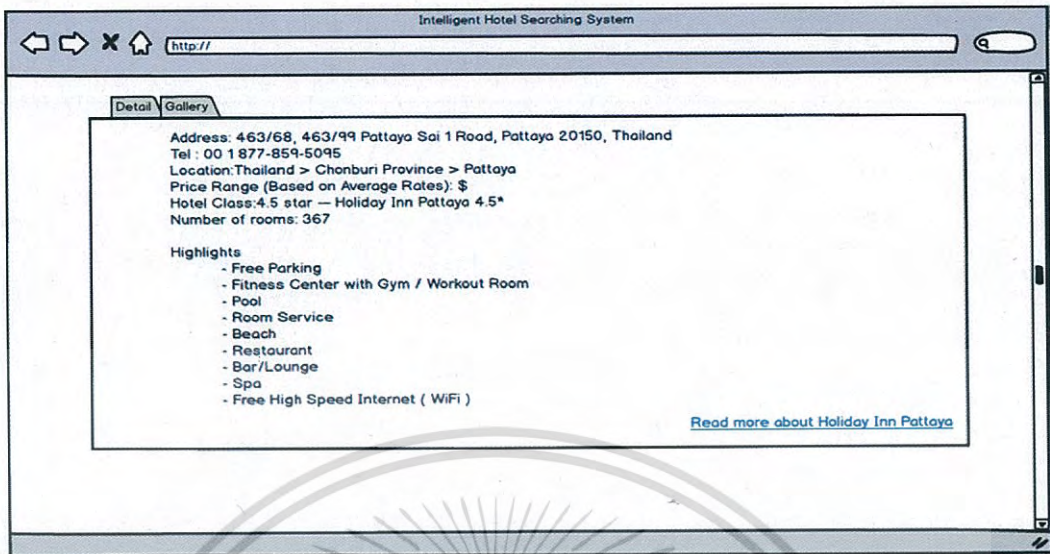
รูปที่ 3.15 หน้าของการค้นหาโรงแรม

หลังจากทำการค้นหาโรงแรมและเลือกโรงแรมที่ต้องการแล้วแสดงได้ดังรูปที่ 3.16 ซึ่งประกอบด้วยหน้าแสดงรายละเอียดของโรงแรมดังรูปที่ 3.17 และหน้าแสดงรูปภาพเพิ่มเติมของโรงแรมแสดงได้ดังรูปที่ 3.18

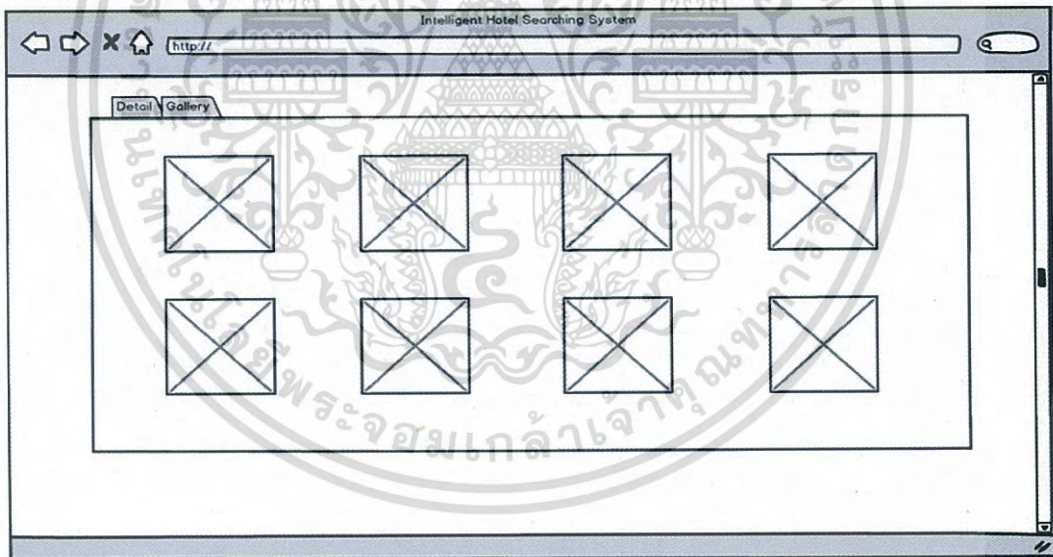


รูปที่ 3.16 หน้าแสดงรายละเอียดของโรงแรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.17 หน้าแสดงรายละเอียดของโรงแรม



รูปที่ 3.18 หน้าแสดงรูปภาพเพิ่มเติมของโรงแรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

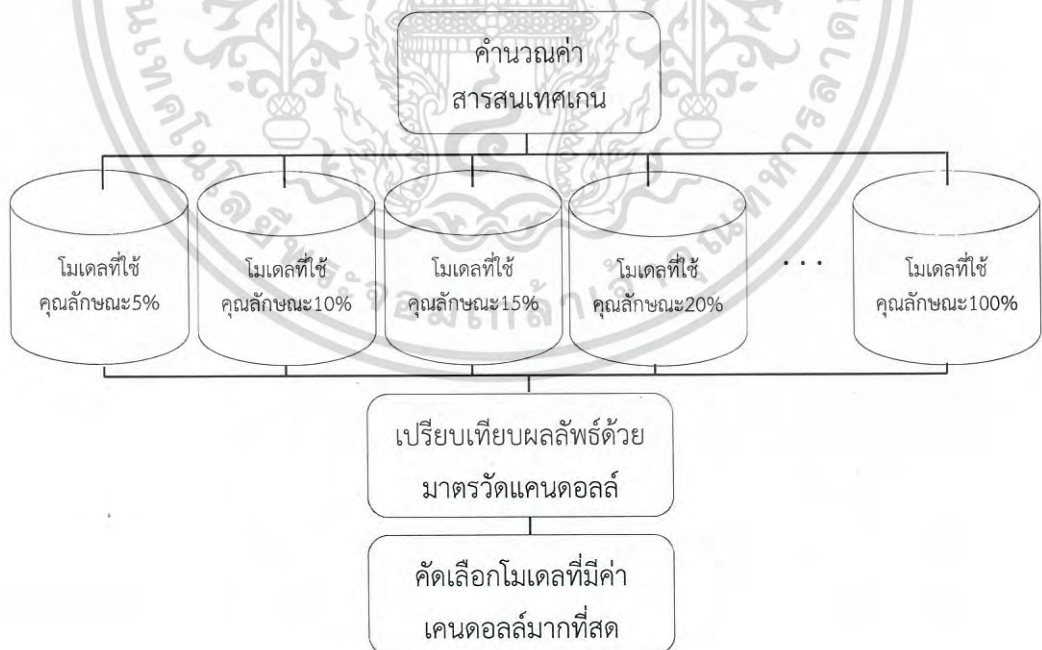
การทดลอง ผลการทดลองและการพัฒนาเว็บแอปพลิเคชัน

ในบทนี้จะกล่าวถึงการทดลอง ผลการทดลอง โดยประยุกต์ใช้ตัวจำแนกประเภทแบบเบย์อย่างง่ายมาสร้างโมเดลเพื่อค้นหาโรงแรมและเปรียบเทียบประสิทธิภาพการค้นหาโรงแรมด้วยวิธีการคัดเลือกคุณลักษณะค่าสารสนเทศเกน (Information Gain) และการพัฒนาเว็บแอปพลิเคชันโดยใช้ PHP เป็นภาษาหลักในการพัฒนา มีการออกแบบหน้าจอแสดงผลให้เรียบง่าย ทันสมัยและสะดวกต่อการใช้งาน

4.1 การทดลองและผลการทดลอง

4.1.1 การทดลอง

ในการทดลองนี้ได้มีการสร้างโมเดลการค้นหาด้วยตัวจำแนกประเภทแบบเบย์อย่างง่ายจากความคิดเห็น 1,800 ความคิดเห็น มีจำนวนคำศัพท์ทั้งหมด 1,650 คำ และใช้วิธีการคัดเลือกคุณลักษณะค่าสารสนเทศเกน (Information Gain) โดยใช้จำนวนคุณลักษณะที่แตกต่างกันเริ่มจาก 5% 10% เพิ่มขึ้นทีละ 5% จนถึง 100% ของจำนวนคุณลักษณะทั้งหมดเพื่อทำการสร้างโมเดล แล้วนำผลลัพธ์การเรียงลำดับของโรงแรมที่ได้จากการค้นหาไปเปรียบเทียบกับผลลัพธ์การเรียงลำดับโดยใช้คะแนนความคิดเห็นจากเว็บไซต์ Tripadvisor ซึ่งการวัดประสิทธิภาพผลลัพธ์นี้จะใช้มาตรวัดของเคนดอลล์ (Kendoll's Rank Correlation) เพื่อวัดประสิทธิภาพในการค้นหาโรงแรมของโมเดล รูปที่ 4.1 แสดงขั้นตอนในการทดลองเพื่อทำการคัดเลือกโมเดลที่ให้ประสิทธิภาพสูงสุด



รูปที่ 4.1 กระบวนการทดลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทดลองนี้ได้กำหนดชุดคำค้นหา (Keywords) เพื่อใช้ในการทดสอบประสิทธิภาพในการค้นหาของโมเดล แสดงได้ดังตารางที่ 4.1

ตารางที่ 4.1 ชุดคำค้นหาเพื่อใช้ในการทดสอบประสิทธิภาพ

หมวดการค้นหา	ชุดคำค้นหา
ห้องพัก	room, clean ,balcony, beautiful, comfort
สถานที่	beach, street, shop, bus, restaurant
ความสะอาด	room, clean, perfect, great
การนอน	bed, quiet, noise, sleep, children
บริการ	breakfast, bar, wifi, gym, pool, spa
ทุกด้าน	clean, room, beach, gym, well, quiet

4.1.2 ผลการทดลอง

ในขั้นตอนนี้จะทำการเปรียบเทียบผลลัพธ์ของแต่ละจำนวนคุณลักษณะโดยใช้ชุดคำค้นหาที่กำหนด และใช้มาตรวัดของเคนดอลล์ในการคัดเลือกโมเดล ซึ่งผลลัพธ์ที่ได้จากการทดลองแสดงได้ดังตารางที่ 4.2 – 4.3

ตารางที่ 4.2 ค่าเคนดอลล์ที่ได้จากโมเดลการคัดเลือกคุณลักษณะด้วยเทคนิคค่าสารสนเทศเกิน 5% ถึง 50%

มาตรวัดของเคนดอลล์	หมวดการค้นหา						ค่าเฉลี่ย
	ค้นหา	สถานที่	ความสะอาด	การนอน	บริการ	ทุกด้าน	
5%	0.019	0.010	-0.002	-0.118	0.072	0.174	0.026
10%	0.014	0.010	0.014	-0.028	-0.002	0.112	0.019
15%	-0.112	0.049	0.068	-0.006	0.058	0.020	0.013
20%	0.092	-0.010	0.093	-0.021	0.025	0.119	0.049
25%	0.041	-0.063	-0.018	-0.030	0.032	0.055	0.003
30%	0.127	0.005	-0.073	0.094	-0.029	0.049	0.029
35%	0.053	0.132	-0.033	-0.001	-0.007	0.081	0.037
40%	0.032	-0.034	-0.024	-0.176	-0.136	0.131	-0.034
45%	-0.082	0.033	0.032	-0.209	0.003	0.165	-0.009
50%	-0.164	0.082	-0.038	-0.229	-0.031	0.096	-0.047

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.3 ค่าเคนดอลล์ที่ได้จากโมเดลการคัดเลือกคุณลักษณะด้วยเทคนิคค่าสารสนเทศเกิน 55% ถึง 100%

มาตรวัดของ เคนดอลล์	หมวดการค้นหา						
	ค้นหา	สถานที่	ความ สะอาด	การนอน	บริการ	ทุกด้าน	ค่าเฉลี่ย
55%	-0.22	0.134	-0.025	-0.087	-0.023	0.067	-0.026
60%	-0.121	0.085	-0.028	-0.092	-0.009	0.049	-0.019
65%	-0.085	0.085	-0.068	-0.129	0.102	0.073	-0.004
70%	-0.138	0.14	-0.055	-0.065	0	0.061	-0.009
75%	-0.066	0.156	-0.059	-0.08	-0.103	0.093	-0.009
80%	-0.066	0.161	-0.151	-0.073	-0.123	0.062	-0.032
85%	-0.172	0.136	-0.085	-0.119	-0.146	-0.077	-0.077
90%	-0.221	0.119	-0.159	-0.132	-0.145	0.062	-0.078
95%	0.008	-0.032	-0.108	-0.133	-0.133	-0.115	-0.101
100%	-0.195	-0.008	0.054	-0.125	-0.112	-0.046	-0.072

จากผลการทดลองในตารางที่ 4.2 – 4.3 แสดงค่าเคนดอลล์ในแต่ละด้านและค่าเฉลี่ยของค่าเคนดอลล์ จากนั้นทำการคัดเลือกโมเดลที่มีค่าเฉลี่ยของค่าเคนดอลล์มากที่สุดและพบว่าค่าเฉลี่ยของค่าเคนดอลล์มากที่สุดที่จำนวนคุณลักษณะ 20% ซึ่งมีค่าศัพท์ 330 คำ เพื่อให้ได้ผลลัพธ์จากการค้นหาโรงแรมด้วยโมเดลมีความใกล้เคียงกับผลลัพธ์ที่เรียงตามคะแนนความคิดเห็นของผู้เข้าพักจริงมากที่สุด

4.2 การพัฒนาเว็บแอปพลิเคชัน

การสร้างเว็บแอปพลิเคชันนี้ใช้ภาษา PHP เป็นภาษาหลักในการพัฒนา โดยใช้ CSS ช่วยในการจัดรูปแบบหน้าจอสวนผู้ติดต่อ

4.2.1 ส่วนของหน้าจอแสดงผล

ในส่วนของหน้าจอแสดงผลจะแบ่งออกเป็น 3 ส่วนคือ ส่วนสำหรับค้นหาโรงแรมประกอบไปด้วยการค้นหาโรงแรมโดยชุดคำค้นหา ช่วงราคา หรือบริการเพิ่มเติม ส่วนที่สองสำหรับดูลำดับโรงแรมจากความคิดเห็นในด้านต่างๆ เพื่อช่วยอำนวยความสะดวกในการค้นหาโรงแรมตามความสนใจของผู้ใช้ และส่วนที่สามสำหรับแสดงรายละเอียด ข้อมูลเพิ่มเติมและรูปภาพของโรงแรม

ส่วนของการค้นหาโรงแรมผู้ใช้สามารถทำการค้นหาได้ในหน้าหลักของระบบอัจฉริยะในการค้นหาโรงแรม โดยภายในหน้าหลักจะทำการแบ่งออกเป็น 4 ส่วนด้วยกัน คือ ส่วนแรกเป็นส่วนแสดงเมนู ส่วนที่สองสำหรับให้ผู้ใช้กรอกชุดคำค้นหาที่ผู้ใช้ต้องการ ส่วนที่สามสำหรับให้ผู้ใช้คลิกเลือกราคา และบริการเพิ่มเติมที่ผู้ใช้สนใจ ส่วนที่สี่เป็นส่วนแสดงโรงแรมที่ระบบทำการสุ่มเลือกขึ้นมาแสดงจำนวน 9 โรงแรม ดังรูปที่ 4.2

Hotel Name	Star Rating	Price (THB)
Eastiny Seven	★★★★☆	935.00
Intimate Hotel	★★★★☆	1,869.00
LK The Empress	★★★★☆	2,179.00
Mer cure Pattaya Ocean Resort	★★★★☆	2,070.00
Sea me Spring Hotel	★★★★☆	1,644.00
The Bayview Hotel	★★★★☆	1,869.00
Cape Dara Resort	★★★★☆	5,470.00
Rabbit Resort	★★★★☆	3,400.00
Sheraton Pattaya Resort	★★★★☆	6,000.00

รูปที่ 4.2 หน้าหลักของระบบอัจฉริยะในการค้นหาโรงแรม











สำหรับการค้นหาโรงแรมผู้ใช้ทำการกรอกชุดคำค้นหา เลือกช่วงราคา เลือกบริการเพิ่มเติม หรือเลือกทำอย่างใดอย่างหนึ่งที่ใช้ต้องการค้นหา ระบบจะทำการค้นหาโรงแรมที่ใกล้เคียงกับที่ผู้ใช้ต้องการมากที่สุดมาแสดง ดังรูปที่ 4.3 แสดงการค้นหาโดยใช้ชุดคำค้นหาว่า clean, room, beach, well, quiet เลือกช่วงราคาระหว่าง 1,500 บาท ถึง 2,500 บาท เลือกบริการเพิ่มเติมเป็น Internet/Wifi, Gym/Fitness

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Home Rank Service

What are you looking for? Search

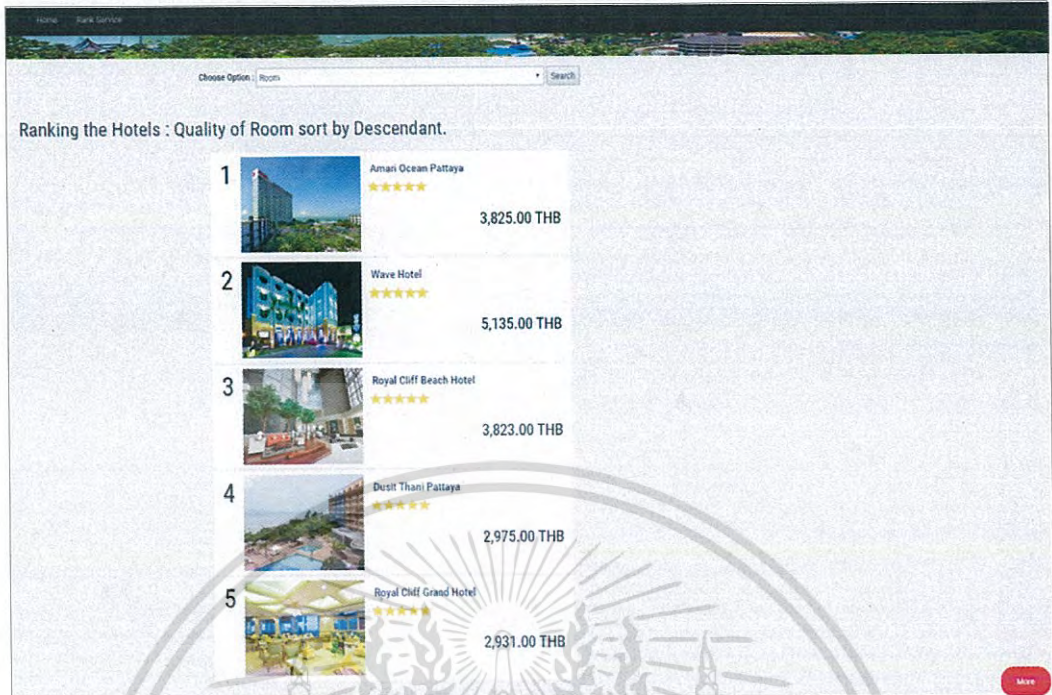
Your recent searches : clean room beach gym well quiet 1500 internet wifi gym fitness

Price	1	2
<input type="radio"/> 500-1500 THB <input type="radio"/> 1500-2500 THB <input type="radio"/> 2500 THB UP	 <p>Bella Villa Prima Hotel (30 Reviews) 1,500.00 THB</p>	 <p>Pinnacle Grand Jomtien Resort (30 Reviews) 1,550.00 THB</p>
Highlights	3	4
<input type="checkbox"/> Pool <input type="checkbox"/> Spa <input type="checkbox"/> Bar/ Lounge <input type="checkbox"/> Internet/ Wifi <input type="checkbox"/> Gym/ Fitness <input type="checkbox"/> Restaurant <input type="checkbox"/> Suite <input type="checkbox"/> Family <input type="checkbox"/> Car Park	 <p>Amari Nova Suites Pattaya (30 Reviews) 2,050.00 THB</p>	 <p>Pattaya Discovery Beach Hotel (30 Reviews) 1,971.00 THB</p>
	 <p>Intimate Hotel (30 Reviews) 1,869.00 THB</p>	 <p>Aya Boutique Hotel (30 Reviews) 1,700.00 THB</p>
	 <p>Grand Jomtien Palace (30 Reviews) 1,534.00 THB</p>	 <p>The Bayview Hotel (30 Reviews) 1,869.00 THB</p>
	 <p>A One The Royal Cruise Hotel (30 Reviews) 1,800.00 THB</p>	 <p>Amari Garden Pattaya (30 Reviews) 1,890.00 THB</p>

รูปที่ 4.3 หน้าจอแสดงผลการค้นหา

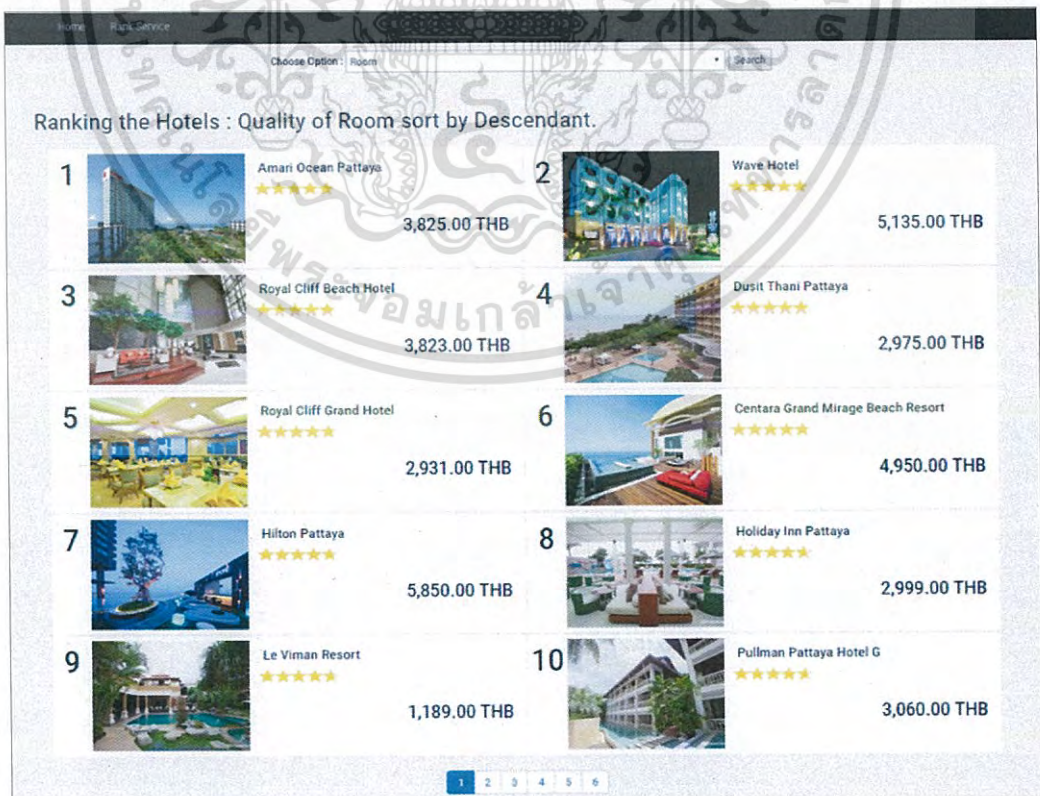
ถัดมาคือส่วนของการดูลำดับโรงแรมจากความคิดเห็นทางด้านต่างๆ ซึ่งระบบจะทำการเรียงลำดับโรงแรมจากคะแนนความคิดเห็นในด้านต่างๆ ประกอบไปด้วยด้านห้องพัก ด้านสถานที่ ด้านความสะอาด ด้านคุณภาพการนอน ด้านบริการ และคะแนนรวมทุกด้าน เมื่อผู้ใช้เลือกด้านที่ผู้ใช้สนใจ ระบบจะแสดงผลการเรียงลำดับ 5 โรงแรมแรกที่มีคะแนนในด้านนั้นๆ สูงที่สุดโดยเรียงลำดับจากมากไปน้อย ดังรูปที่ 4.4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.4 หน้าจอแสดงผลการเรียงลำดับด้านห้องพัก 5 โรงแรมแรก

และสามารถดูผลการเรียงลำดับของโรงแรมทั้งหมด 60 โรงแรมในด้านนั้นๆ ได้ด้วยการคลิกปุ่มเพิ่มเติม (More) ที่มุมขวาล่างของหน้าจอ รูปที่ 4.5 หน้าจอแสดงผลการเรียงลำดับด้านห้องพักทั้งหมด 60 โรงแรม



รูปที่ 4.5 หน้าจอแสดงผลการเรียงลำดับด้านห้องพักทั้งหมด 60 โรงแรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนสุดท้ายคือส่วนของการแสดงรายละเอียด ข้อมูลเพิ่มเติมและรูปภาพของโรงแรม ซึ่งรายละเอียดของโรงแรมจะประกอบไปด้วยชื่อโรงแรม ที่อยู่และสามารถคลิกดูรายละเอียดแบบเต็มได้ โดยจะแสดงคำอธิบายเกี่ยวกับโรงแรม จำนวนห้องพักที่มี และลิงค์ไปยังเว็บไซต์ Tripadvisor สำหรับดูข้อมูลเพิ่มเติมของโรงแรม ดังรูปที่ 4.6 – 4.7



รูปที่ 4.6 หน้าจอแสดงรายละเอียดของโรงแรม



รูปที่ 4.7 หน้าจอแสดงรายละเอียดแบบเต็มของโรงแรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในหน้าจอบรรยากาศของรีสอร์ท จะแสดงบรรยากาศของรีสอร์ทโดยจะทำการเลื่อนไปยังรูปถัดไปโดยอัตโนมัติ หากผู้ใช้สนใจรูปภาพใดเป็นพิเศษสามารถคลิกดูได้จากรูปเล็กๆที่อยู่ด้านล่างของหน้าจอ ดังรูปที่ 4.8



รูปที่ 4.8 หน้าจอบรรยากาศของรีสอร์ท

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทสรุปและข้อเสนอแนะ

5.1 บทสรุป

ระบบอัจฉริยะในการค้นหาโรงแรมเป็นเว็บแอปพลิเคชันที่พัฒนาขึ้นเพื่ออำนวยความสะดวกในการค้นหาโรงแรมให้ตรงตามความต้องการของผู้ใช้มากที่สุดโดยนำข้อมูลความคิดเห็นของผู้ใช้มาช่วยในการประมวลผลและประยุกต์ใช้เทคนิคการทำเหมืองข้อมูล การคัดเลือกคุณลักษณะของข้อมูล การจำแนกประเภทแบบเบย์อย่างง่าย โดยขั้นตอนการสร้างโมเดลเพื่อใช้ในการค้นหาเริ่มจากการเก็บความคิดเห็นที่มีอยู่บนเว็บไซต์ แล้วทำการนับความถี่ในการปรากฏของคำศัพท์จากความคิดเห็นลำดับถัดมาทำการหาค่าความน่าจะเป็นด้วยตัวจำแนกประเภทแบบเบย์อย่างง่าย หลังจากนั้นนำโมเดลที่ได้ไปทดสอบผลการค้นหาจากชุดคำค้นหา โดยเปรียบเทียบการค้นหาซึ่งใช้มาตรวัดความสอดคล้องของเคนดอลล์เป็นตัวชี้วัดประสิทธิภาพการค้นหาโรงแรม และทำการคัดเลือกโมเดลที่มีค่าเคนดอลล์ที่ดีที่สุด

ระบบอัจฉริยะในการค้นหาโรงแรมมีความสามารถในการค้นหาโรงแรมโดยการกรอกชุดคำค้นหา เลือกช่วงราคาหรือบริการเพิ่มเติมที่ผู้ใช้งานต้องการค้นหา และสามารถดูลำดับโรงแรมจากคะแนนความคิดเห็นในด้านต่างๆ โดยเรียงจากคะแนนที่ผู้เข้าพักได้ให้ไว้ต่อโรงแรมจากมากไปน้อย นอกจากนี้โมเดลการค้นหายังสามารถเรียนรู้ออนไลน์ได้อัตโนมัติเมื่อมีผู้ใช้ใหม่ได้แสดงความคิดเห็นต่อโรงแรม ระบบจะทำการนับความถี่ในการปรากฏคำศัพท์และดึงคะแนนที่ผู้เข้าพักได้ให้ไว้ต่อโรงแรมโดยอัตโนมัติ

5.2 ข้อเสนอแนะ

- ค่าเคนดอลล์ที่คำนวณออกมาไม่เป็นที่น่าพอใจ เนื่องจากมีจำนวนโรงแรมที่ต้องจำแนกประเภทมากถึง 60 โรงแรม แต่มีจำนวนความคิดเห็นเพียงโรงแรมละ 30 ความคิดเห็นเท่านั้น ดังนั้นควรมีจำนวนความคิดเห็น 100 ความคิดเห็นต่อ 1 โรงแรมเพื่อให้มีสัดส่วนของจำนวนตัวอย่างสอดคล้องกับจำนวนโรงแรม
- เนื่องจากระบบที่ทำการตัดคำแต่ละคำ เพื่อนำไปคิดค่าความน่าจะเป็น เช่น คำว่า “internet” หมายถึง โรงแรมนี้มีอินเทอร์เน็ต แต่ถ้าคำศัพท์ที่ผู้ใช้งานต้องการค้นหาเป็นคำว่า “no internet” ซึ่งเป็นความคิดเห็นในทางลบ ที่หมายความว่าโรงแรมนี้ไม่มีอินเทอร์เน็ต อาจทำให้เกิดความผิดพลาดในการค้นหา ดังนั้นจึงควรมีการพัฒนาโมเดลการค้นหาโรงแรมที่มีแนวคิดในการหาความสัมพันธ์ระหว่างคำศัพท์ได้ เพื่อให้การค้นหาโรงแรมมีประสิทธิภาพมากยิ่งขึ้น
- เนื่องจากการดึงความคิดเห็นเป็นการดึงทั้งหน้าของเว็บไซต์โดยอัตโนมัติด้วยเอเจนต์ของระบบที่พัฒนาขึ้น จึงทำให้ไม่ได้คัดเลือกความคิดเห็นที่เหมาะสม บางความคิดเห็นอาจมีคะแนนที่ผู้ใช้ให้กับโรงแรมไม่สอดคล้องกับข้อความที่ผู้ใช้แสดงความคิดเห็น ดังนั้นควรเลือกความคิดเห็นที่มีเนื้อความสอดคล้องกับคะแนนที่ผู้ใช้ให้ เพื่อให้การค้นหาโรงแรมมีความถูกต้องมากยิ่งขึ้น

เอกสารอ้างอิง

- [1] นิเวศ จิระวิจิตชัย. 2553. การค้นหาเทคนิคเหมืองข้อมูลเพื่อสร้างโมเดลการวิเคราะห์โรคอัตโนมัติ. มหาวิทยาลัยขอนแก่น.
- [2] อังศุมาลี สุทธิภักดี. 2553. การเปรียบเทียบประสิทธิภาพของการจำแนกหมวดหมู่ของข้อคิดเห็นในแบบสอบถามแบบเปิด โดยใช้วิธีนาอิวเบย์และซัพพอร์ตเวกเตอร์แมชชีน. วิทยาลัยราชพฤกษ์.
- [3] จันสธา ศรีสรवल. 2557. “การค้นหาโรงแรมด้วยเทคนิคการรวมกลุ่มแบบเบย์อย่างง่าย.” วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.
- [4] Zafarani, R. Ali, M. and Liu, H. 2014. *Social Media Mining: An Introduction*. Draft Version. Cambridge University Press.
- [5] Troussas, C. Virvou, M. Espinosa, K. J. Llaguno, K. and Caro, J. 2013. “Sentiment Analysis of Facebook Status Using Naïve Bayes Classifier for Language Learning.” In *proceeding of Fourth International Conference on Information Intelligence Systems and Applications (IISA)*.
- [6] Ting, S.L. Ip, W.H. Tsang, Albert H.C. 2011. “Is Naïve Bayes a Good Classifier for Document Classification?.” *International Journal of Software Engineering and Its Applications*. 5(3). 37-46.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก



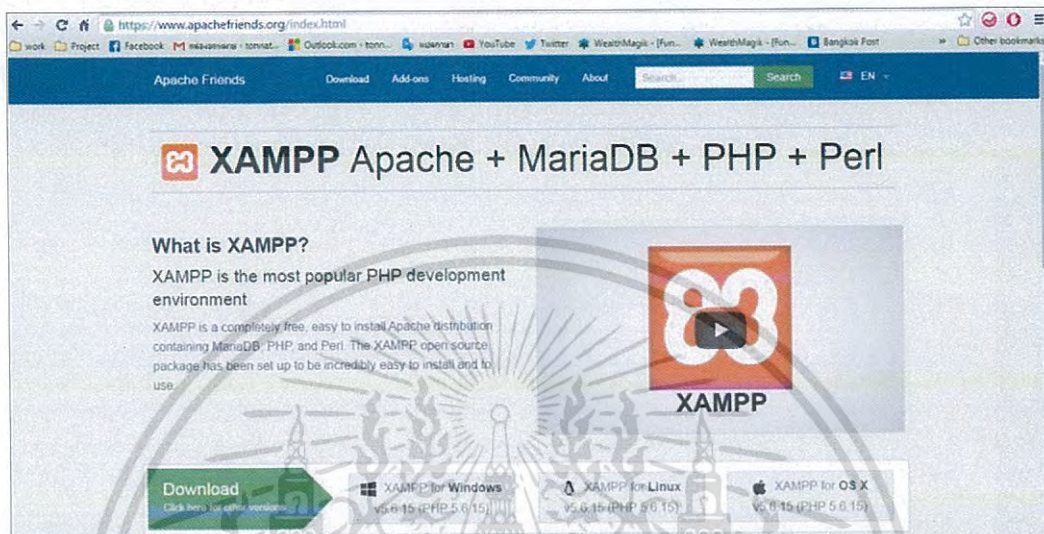
ขั้นตอนการติดตั้งระบบอัจฉริยะในการค้นหาโรงแรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนการติดตั้งระบบอัจฉริยะในการค้นหาโรงแรม

1. ติดตั้งโปรแกรม XAMPP

โปรแกรม XAMPP ใช้สำหรับจำลองเครื่องคอมพิวเตอร์ส่วนบุคคลให้ทำงานในลักษณะเว็บเซิร์ฟเวอร์ ดาวน์โหลด XAMPP จากเว็บไซต์ www.apachefriends.org



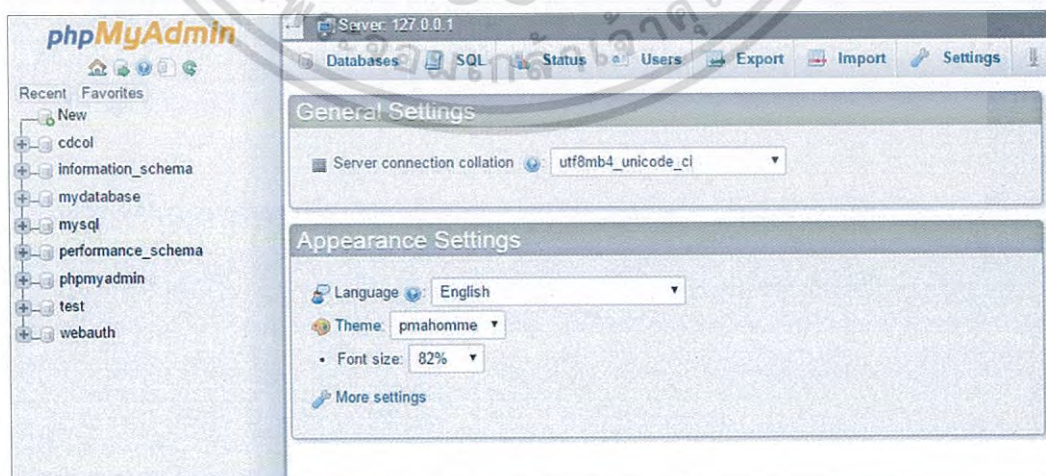
รูปที่ ก.1 เว็บไซต์ www.apachefriends.org

2. ติดตั้ง laravel framework

Laravel คือ PHP Framework ที่ช่วยให้ง่ายต่อการพัฒนาเว็บแอปพลิเคชัน สามารถอ่านและทำความเข้าใจโค้ดได้ง่าย ออกแบบมาเพื่อพัฒนาเว็บแอปพลิเคชันในรูปแบบ *Model-view-controller* (MVC)

3. คัดลอกเพิ่ม project ไปยังไดเรกทอรี C:\xampp\htdocs

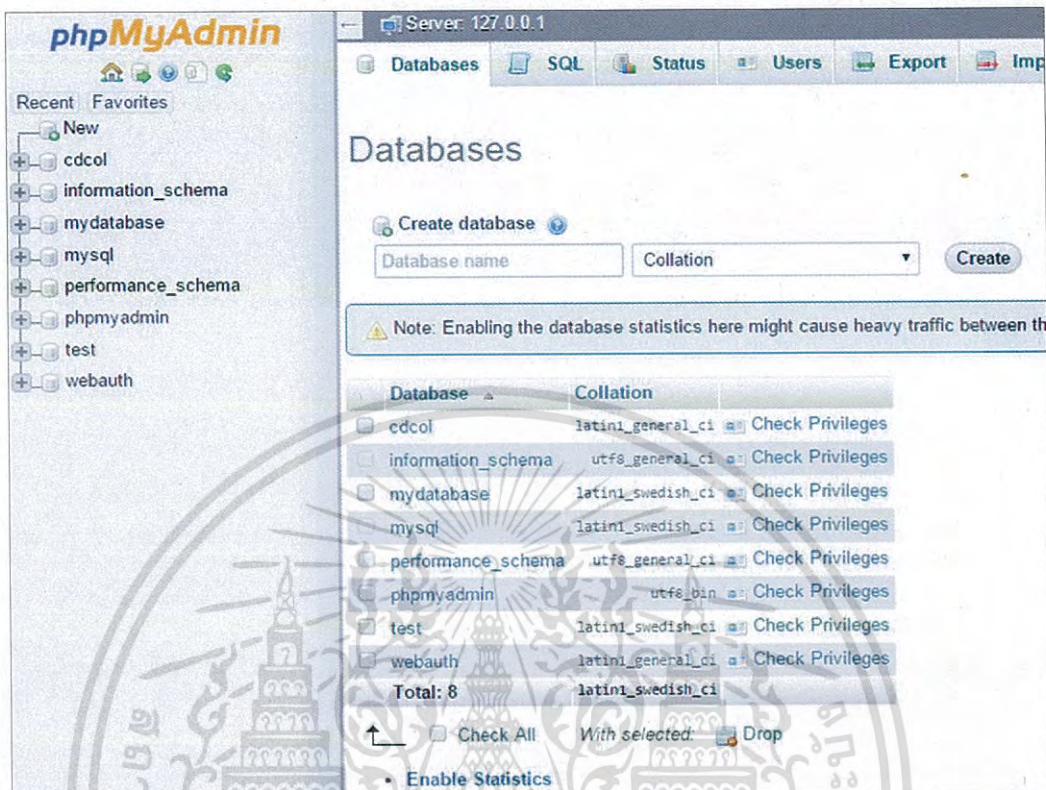
4. เปิดเบราว์เซอร์ ใส่ URL เป็น localhost/phpmyadmin



รูปที่ ก.2 หน้า phpMyadmin

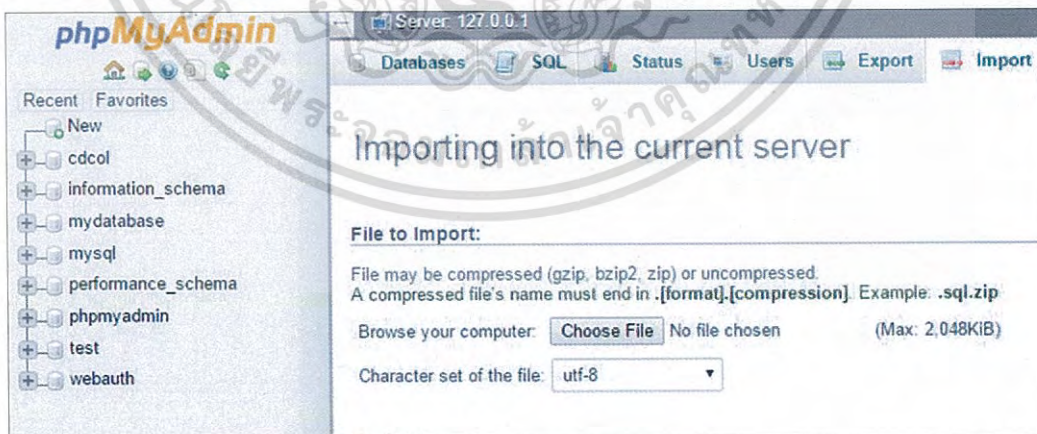
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. คลิก New  แล้วสร้างฐานข้อมูล



รูปที่ ก.3 หน้าสร้างฐานข้อมูล

6. เมื่อสร้างฐานข้อมูลเสร็จแล้วคลิกปุ่ม Import แล้วเลือกไฟล์ data.sql จากนั้นคลิกปุ่ม Go



รูปที่ ก.4 หน้า Import ไฟล์ SQL

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีคนนำไปใช้

7. ทดลองเข้าเว็บแอปพลิเคชัน ใส่ URL เป็น localhost/project/public จะแสดงหน้าเว็บเพจดังรูปด้านล่าง



รูปที่ ก.5 หน้าเว็บเพจระบบอัจฉริยะในการค้นหาโรงแรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

โค้ดโปรแกรมของเอเจนต์ที่ใช้ในระบบอัจฉริยะในการค้นหา
โรงแรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนการสร้างเอเจนต์ ในการดึงข้อมูลจากเว็บไซต์

เอเจนต์ ในระบบมี 2 ไฟล์ ไฟล์แรกชื่อ feedid.php ทำหน้าที่ในการดึงไอดี (ID) ของความคิดเห็น ไฟล์ที่สองชื่อ feed.php ทำหน้าที่ในการดึงข้อมูลความคิดเห็น โดยทั้ง 2 ไฟล์จะต้องดาวน์โหลดไลบรารี (library) ที่ <http://sourceforge.net/projects/simplehtmldom/files/> ก่อน และนำไฟล์ php ชื่อว่า simple_html_dom.php ที่ได้มาไว้ที่โฟลเดอร์ที่เก็บรวมไฟล์ HTML หลังจากนั้นเขียนโค้ด php โดยภายในโค้ดของไฟล์นั้นจะต้องนำเข้าไปไฟล์ simple_html_dom.php ด้วย

โค้ดการดึงข้อมูลจากหน้าเว็บเพจ

```
if ($data2["status_update"] == "allow") {
    $sql = "SELECT hotel_id as ID,link AS LINK
          FROM hotels";
    $result = $conn->query($sql);

    while($row = $result->fetch_assoc()) {
        $target = file_get_html($row['LINK']);
        echo $row['LINK']."<br>";
        $item = array();

        foreach($target->find('div#REVIEWS div.reviewSelector') as $sel)
        {
            $data["id_update"][$row['ID']-1]["id_review"][] = str_replace("review_",
            "r", $sel->id);
            $item[] = str_replace("review_", "r", $sel->id);
        }

        foreach ($item as $value) {
            echo $value."<br>";
        }
        echo "<hr>";
    }
    $data["status_update"] = "allow";
    $data2["status_update"] = "no";
}
```

รูปที่ ข.1 การดึงไอดีของความคิดเห็น

จากรูปที่ ข.1 เป็นการเช็คสถานะการอนุญาตให้ดึงไอดี (ID) ของความคิดเห็น ถ้าสถานะของ status_update ในไฟล์ idupdate.json เป็น allow แสดงว่าอนุญาตให้ดึงไอดีเข้ามาใหม่ได้ ซึ่งจะทำให้การดึง html code มาจากหน้าเว็บเพจที่ระบุจาก URL จากนั้นจึงค้นหาไอดีของความคิดเห็นตาม path ที่กำหนดไว้ โดยวนซ้ำจนครบทุกโรงแรม หลังจากนั้นจะทำการกำหนดสถานะของ status_update ในไฟล์ idupdate.json ให้เป็น no และของ status_update ในไฟล์ update.json ให้เป็น allow เพื่อป้องกันการดึงไอดีเข้ามาแทนที่ไอดีเก่าในกรณีที่ไม่สามารถทำการดึงข้อมูลความคิดเห็นได้สมบูรณ์

ในส่วนของโค้ด รูปที่ ข.2 เป็นการเช็คสถานะการอนุญาตให้ดึงข้อมูลความคิดเห็น ถ้าสถานะของ status_update ในไฟล์ update.json เป็น allow แสดงว่าอนุญาตให้ดึงข้อมูลความคิดเห็นได้ ในการดึงข้อมูลแบ่งออกเป็น 3 ส่วน

ส่วนแรกจะดึงคะแนนความคิดเห็นโดยรวมของความคิดเห็นนั้น จากนั้นจะทำการกำหนดคะแนนความคิดเห็นในแต่ละด้านให้มีค่าเท่ากับคะแนนความคิดเห็นโดยรวม

```

if ($data["status_update"] == "allow") {
    foreach ($data["id_update"] as $value) {
        $data2["review_update"][$i]["old"] = $data2["review_update"][$i]["new"];
        $text = $value["id_hotel"];
        foreach ($value["id_review"] as $id) {
            if ($id == $data2["review_update"][$i]["new"]) {
                break;
            } else {
                $domain = str_replace("-Reviews-", "-.$id.-", $text);
                $target = file_get_html($domain);
                $array = array();
                foreach($target->find('div#'.str_replace("r", "review_", $id).
                    ' div.rating span.rate img') as $sel)
                {
                    $array["All"] = substr($sel->alt, 0, strpos($sel->alt, " "));
                    $array["Rooms"] = substr($sel->alt, 0, strpos($sel->alt, " "));
                    $array["Service"] = substr($sel->alt, 0, strpos($sel->alt, " "));
                    $array["Location"] = substr($sel->alt, 0, strpos($sel->alt, " "));
                    $array["Cleanliness"] = substr($sel->alt, 0, strpos($sel->alt, " "));
                    $array["SleepQuality"] = substr($sel->alt, 0, strpos($sel->alt, " "));
                }
            }
        }
    }
}

```

รูปที่ ข.2 การดึงคะแนนความคิดเห็นโดยรวมของความคิดเห็น

ส่วนที่สองเป็นการดึงคะแนนความคิดเห็นในแต่ละด้านของความคิดเห็นนั้น ประกอบไปด้วย ด้านห้องพัก สถานที่ ความสะอาด คุณภาพการนอนและบริการมาเก็บไว้ในอาร์เรย์ ดังรูปที่ ข.3

```

foreach($target->find('div#'.str_replace("r", "review_", $id).' div.rating-list
li.recommend-answer') as $sel)
{
    preg_match('/alt="(.*?)"/', $sel, $matches);
    if(str_replace(" ", "", $sel->plaintext) == "Rooms") {
        $array["Rooms"] = substr($matches[1], 0, strpos($matches[1], " "));
    }
    if(str_replace(" ", "", $sel->plaintext) == "Service") {
        $array["Service"] = substr($matches[1], 0, strpos($matches[1], " "));
    }
    if(str_replace(" ", "", $sel->plaintext) == "Location") {
        $array["Location"] = substr($matches[1], 0, strpos($matches[1], " "));
    }
    if(str_replace(" ", "", $sel->plaintext) == "Cleanliness") {
        $array["Cleanliness"] = substr($matches[1], 0, strpos($matches[1], " "));
    }
    if(str_replace(" ", "", $sel->plaintext) == "SleepQuality") {
        $array["SleepQuality"] = substr($matches[1], 0, strpos($matches[1], " "));
    }
}
print_r($array);
echo "<br>";
if(!empty($array)) {
    $result[] = array(
        "id" => $id,
        "room" => $array["Rooms"],
        "locate" => $array["Location"],
        "clean" => $array["Cleanliness"],
        "sleep" => $array["SleepQuality"],
        "service" => $array["Service"],
        "all" => $array["All"],
        "idHotel" => $data2["review_update"][$i]["id_hotel"],
    );
}

```

รูปที่ ข.3 การดึงคะแนนความคิดเห็นในแต่ละด้านของความคิดเห็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนที่สามเป็นการดึงความคิดเห็นมาทำการลบตัวเลข สัญลักษณ์และอักขระพิเศษออกจากความคิดเห็น จากนั้นนำความคิดเห็นมาตัดแบ่งเป็นแต่ละคำ แล้วทำการเปรียบเทียบกับคำศัพท์ที่มีอยู่ในพจนานุกรมของแต่ละโรงแรม ถ้ามีคำศัพท์ตรงกันจะเพิ่มความถี่ของคำศัพท์นั้นในพจนานุกรม ดังรูปที่ ข.4

```
foreach($target->find('div#'.str_replace("r", "review_", $id).' div.entry p') as $el)
{
    $words = explode(" ", str_replace($vowels, "", strtolower($el->plaintext)));
}

for ($j=0; $j < count($words); $j++) {
    if(empty($words[$j])) {
    } else {
        for ($k=$alpha[substr($words[$j], 0, 1)]["start"]; $k <=
        $alpha[substr($words[$j], 0, 1)]["end"]; $k++) {
            if($models[$i][$k]["words"] == $words[$j]) {
                $models[$i][$k]["freq"]++;
            }
        }
    }
}
```

รูปที่ ข.4 การดึงความคิดเห็น

หลังจากทำการปรับปรุงข้อมูลในฐานข้อมูลสำเร็จ จะทำการกำหนดสถานะของ status_update ในไฟล์ update.json ให้เป็น no และของ status_update ในไฟล์ idupdate.json ให้เป็น allow ดังรูปที่ ข.5

```
$data["status_update"] = "no";
$data2["status_update"] = "allow";
```

รูปที่ ข.5 การกำหนดสถานะ