

รหัสโครงการ 13P33C001

ระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรม
BEST 2011: การแข่งขันสุดยอดซอฟต์แวร์ประมวลผลภาษาไทย

รายงานฉบับสมบูรณ์

เสนอต่อ

ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ

กระทรวงวิทยาศาสตร์และเทคโนโลยี

ได้รับทุนอุดหนุนโครงการวิจัย พัฒนาและวิศวกรรม
โครงการแข่งขันพัฒนาโปรแกรมคอมพิวเตอร์แห่งประเทศไทย ครั้งที่ 13
ประจำปีงบประมาณ 2553

โดย

น.ส.วรัญญา วรรณศรี

อาจารย์ ดร.เด่นดวง ประดับสุวรรณ

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

กิตติกรรมประกาศ

การจัดทำระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรม (Opinion Mining on Hotel Reviews) สามารถก้าวมาถึงจุดนี้ เนื่องจากได้รับคำแนะนำ และความช่วยเหลือต่างๆ จากบุคคลเหล่านี้มา โดยตลอด ผู้พัฒนาจึงขอขอบพระคุณบุคคลดังต่อไปนี้

อาจารย์ ดร.เด่นดวง ประดับสุวรรณ อาจารย์ประจำภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ซึ่งเป็นอาจารย์ที่ปรึกษาในโครงการนี้ ได้คอยให้คำแนะนำ ปรึกษาและชี้แนะแนวทางในการดำเนินงานที่เป็นประโยชน์ให้ลุล่วงไปได้ด้วยดี

ดร.ชูชาติ หฤไชยะศักดิ์ หัวหน้างานเทคโนโลยีโครงสร้างพื้นฐานสารสนเทศอัจฉริยะ หน่วยปฏิบัติการวิจัยวิทยาการมนุษยภาษา ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ที่คอยให้คำปรึกษาในด้านการวิเคราะห์และสกัดข้อความแสดงความคิดเห็น ชี้แนะแนวทางในการพัฒนาโครงการในด้านการวิจัยและเทคโนโลยีต่างๆที่ใช้ในการพัฒนา

และสุดท้ายนี้ขอขอบคุณโครงการการแข่งขันพัฒนาโปรแกรมคอมพิวเตอร์แห่งประเทศไทยครั้งที่ 13 ซึ่งจัดขึ้นโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ ที่มอบทุนอุดหนุนการพัฒนาบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรม ทำให้การพัฒนาสามารถผ่านลุล่วงมาได้ด้วยดี

นางสาวรัญญา วรรณศรี

หัวหน้าโครงการ

บทคัดย่อ

เทคโนโลยีเหมืองข้อความแสดงความคิดเห็นเป็นเทคโนโลยีที่ผสมผสานเทคนิคของการสืบค้นข้อมูลเข้ากับการประมวลผลทางภาษา โดยเน้นการวิเคราะห์ข้อความแสดงความคิดเห็นต่างๆเพื่อนำมาใช้ประโยชน์ในการวางแผนต่อไป จึงเป็นอีกหนึ่งเทคโนโลยีที่เข้ามาช่วยในการประเมินความพึงพอใจของลูกค้าที่มีต่อสินค้าหรือการให้บริการ และทางผู้พัฒนาได้เล็งเห็นว่า ประเทศไทยเป็นประเทศท่องเที่ยวที่มีชื่อเสียง มีที่พักมากมายสำหรับรองรับนักท่องเที่ยว การนำข้อความที่อยู่ในรูปแบบการแสดงความคิดเห็นด้วยภาษาธรรมชาติของผู้ที่เคยพักในโรงแรมต่างๆมาวิเคราะห์ข้อมูล โดยอาศัยหลักการการวิเคราะห์และสรุปผลจากข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรมในระดับคุณลักษณะ ซึ่งจะทำให้ผู้ที่ต้องการเลือกพักโรงแรมที่ตรงกับความต้องการ สามารถตัดสินใจได้ง่ายขึ้น และไม่ต้องเสียเวลาไปกับการอ่านข้อความแสดงความคิดเห็นจำนวนมาก ดังนั้น ระบบนี้จึงถูกพัฒนาขึ้นโดยมีวัตถุประสงค์ในการวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม โดยระบบจะรับข้อความแสดงความคิดเห็นภาษาไทย มาวิเคราะห์แล้วระบุว่าข้อความแสดงความคิดเห็นนั้น แสดงทัศนคติต่อคุณลักษณะใดของโรงแรม จากนั้นจะระบุว่าข้อความแสดงความคิดเห็นนั้น มีข้อความความคิดเห็นเป็นบวก เป็นลบ หรือเป็นกลาง แล้วนำข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรมมาสรุปในเชิงสถิติและแสดงผลในรูปแบบกราฟิก

คำสำคัญ: เทคโนโลยีเหมืองข้อความแสดงความคิดเห็น, การวิเคราะห์ข้อความแสดงความคิดเห็น, ข้อความแสดงความคิดเห็นภาษาไทย

Abstract

Thailand is famous tourism country where has a large number of hotels and resorts to support many tourists. If tourist's opinions were analyzed by some techniques which determine their opinions from text written in natural language and summarize the viewpoint of their opinions, Tourists who want to book hotel or resort for their holiday might make their decision easier when they view those summarized opinions.

Opinion mining or Sentiment analysis is the technique that applies text mining technique on human opinion written in natural language. "Opinion Mining System on Hotel Reviews" uses the opinion mining technique with information visualization to create summarized opinion in graphic view. First, the system receives input which is tourist's opinion that is written in Thai language. Then, the opinion will be analyzed by "Feature-based sentiment analysis and summarization" method. The result of this process is the summarized opinion that opinion segments were determined feature and polarity. Finally, summarized opinion was converted to graphic view.

Keywords: Opinion mining, Sentiment analysis, Thai language

สารบัญ

	หน้า
บทที่ 1 บทนำ	1
- ความเป็นมาและสาระสำคัญของโครงการ	1
บทที่ 2 วัตถุประสงค์และเป้าหมายของโครงการ	3
บทที่ 3 รายละเอียดของการพัฒนา	4
- เนื้อเรื่องย่อ	4
- เทคนิคและเทคโนโลยี	6
- เครื่องมือที่ใช้ในการพัฒนา	14
- ขอบเขตและข้อจำกัด	21
บทที่ 4 กลุ่มผู้ใช้ระบบ	22
บทที่ 5 ผลของการทดสอบระบบ	23
บทที่ 6 ปัญหาและอุปสรรค	25
บทที่ 7 แนวทางในการพัฒนาและประยุกต์ใช้ร่วมกับงานอื่นๆในขั้นต่อไป	26
บทที่ 8 ข้อสรุปและข้อเสนอแนะ	27
บทที่ 9 เอกสารอ้างอิง	28
บทที่ 10 ประวัติส่วนตัวและผลงานดีเด่นของผู้พัฒนา	29
บทที่ 11 ภาคผนวก	30
- คู่มือการติดตั้ง	30
- คู่มือการใช้งาน	34

บทที่ 1 บทนำ

ความเป็นมาและสาระสำคัญของโครงการ

อุตสาหกรรมการท่องเที่ยวในประเทศไทยเติบโตขึ้นอย่างมากในทศวรรษที่ผ่านมา นับเป็นอีกด้านหนึ่งที่สร้างชื่อเสียงและรายได้ให้แก่ประเทศเป็นอย่างมาก เนื่องจากประเทศไทยมีทรัพยากรธรรมชาติที่สวยงาม มีเมืองใหญ่ที่เป็นศูนย์รวมความเจริญในหลายๆด้าน สามารถเป็นแหล่งท่องเที่ยวที่เป็นที่นิยมสูงสุดในภูมิภาคนี้ จากมุมมองของนักท่องเที่ยวต่างชาติซึ่งมองว่าประเทศไทยเป็นแหล่งท่องเที่ยวที่ให้ ความอบอุ่น มีมิตรภาพต่อนักท่องเที่ยว ที่พักอาศัยอบอุ่น และนอกจากนี้ปัจจัยที่ช่วยสนับสนุนให้ประเทศไทยได้เปรียบกว่าประเทศอื่นๆในอาเซียนซึ่งนับเป็นคู่แข่งที่สำคัญในอุตสาหกรรมการท่องเที่ยว ได้แก่ ความสามารถในการรองรับด้านที่พักซึ่งมีจำนวนที่พักมากกว่าประเทศอื่นๆในอาเซียน จากบัญชีรายชื่อโรงแรม ตามพระราชบัญญัติโรงแรม พ.ศ.2547 ประจำปี 2550 ประเทศไทยมีโรงแรมสำหรับรองรับนักท่องเที่ยวประมาณ 4,800 แห่ง[10] นับว่าเพียงพอต่อการรองรับนักท่องเที่ยว ดังนั้นหากมีระบบการจัดการที่ดีแล้ว ประเทศไทยจะยังคงความเป็นแหล่งท่องเที่ยวที่เป็นที่นิยมที่สุดในอาเซียน

เมื่อมีการวางแผนการท่องเที่ยว สิ่งสำคัญลำดับต้นๆที่นักท่องเที่ยวต้องสนใจคือ ที่พัก นักท่องเที่ยวจึงมักจะมีการหาข้อมูลล่วงหน้า เพื่อให้ได้ที่พักที่ตรงตามความต้องการมากที่สุด การหาข้อมูลทางอินเทอร์เน็ตนับเป็นอีกวิธีหนึ่งที่สะดวกในการหาข้อมูล เว็บไซต์ในปัจจุบันมักจะมีนักท่องเที่ยวมาแสดงความคิดเห็นของตนเกี่ยวกับสถานที่ต่างๆที่เคยไปมา ให้ข้อมูลต่างๆเพื่อแนะนำผู้ที่จะไปสถานที่นั้นในภายหลัง ซึ่งที่พักบางแห่ง อาจมีผู้มาเขียนแสดงความคิดเห็นไว้เป็นจำนวนมาก อีกทั้งความคิดเห็นของแต่ละบุคคลก็ให้ความหลากหลายในแต่ละมุมมอง เงื่อนไขการตัดสินใจเลือกที่พักของแต่ละบุคคลก็แตกต่างกัน นักท่องเที่ยวบางคนอาจสนใจในเรื่องของการบริการ บางคนอาจสนใจเรื่องอาหาร บางคนอาจต้องการโรงแรมที่มีบรรยากาศเงียบสงบเหมาะแก่การพักผ่อน ซึ่งการที่จะหาความคิดเห็นที่เป็นมุมมองที่ต้องการจากปริมาณความคิดเห็นที่มีเป็นจำนวนมาก อาจต้องใช้เวลา [1] การนำเสนอข้อมูลในระดับคุณลักษณะ (Feature Level) จะช่วยให้ค้นหามุมมองที่ต้องการได้สะดวกมากกว่าการอ่านจากข้อความความคิดเห็นที่ถูกรวบรวมเอาไว้ และเมื่อนำข้อมูลมาสรุปให้อยู่ในรูปแบบกราฟิกจะเป็นอีกหนึ่งทางที่ช่วยให้ทราบถึงแนวทางของความคิดเห็นทั้งหมดได้โดยไม่ต้องอ่านข้อความความคิดเห็นทั้งหมดด้วยตนเอง และช่วยให้สามารถตัดสินใจได้สะดวกและตรงตามความต้องการมากยิ่งขึ้น

ระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรม เป็นระบบที่สามารถวิเคราะห์ข้อมูลที่อยู่ในรูปแบบการแสดงความคิดเห็นด้วยภาษาธรรมชาติ โดยอาศัยหลักการการวิเคราะห์และสรุปผลจากทัศนคติในระดับคุณลักษณะ (Feature-based sentiment analysis and summarization) ทำให้ข้อมูลการแสดงความคิดเห็นอยู่ในรูปแบบที่เป็นข้อมูลสรุปให้เข้าใจได้ง่ายในรูปแบบกราฟิก โดยสามารถเลือกกลับไปดูข้อความแสดงความคิดเห็นที่ถูกนำมาวิเคราะห์ในหัวข้อของมุมมองที่สนใจได้

บทที่ 2 วัตถุประสงค์และเป้าหมายของโครงการ

ศึกษาและวิจัยวิธีการสกัดคำบอกคุณลักษณะและคำระบุชี้แจงแสดงความคิดเห็น(Polar Word) อัตโนมัตินอกจากคลังข้อความแสดงความคิดเห็นสำหรับโรงแรม รวมถึงออกแบบและพัฒนาระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรมโดยสามารถแสดงผลเชิงสถิติและในรูปแบบกราฟิก เพื่อเพิ่มความสะดวกให้แก่นักท่องเที่ยวสำหรับค้นหาข้อมูลความคิดเห็นเกี่ยวกับโรงแรมทางอินเทอร์เน็ตในการเลือกโรงแรมที่ตรงตามความต้องการของตนเอง และลดระยะเวลาในการอ่านข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม ด้วยข้อมูลสรุปในรูปแบบกราฟิก

บทที่ 3 รายละเอียดของการพัฒนา

3.1 เนื้อเรื่องย่อ


นักท่องเที่ยวชื่อ นายซี ต้องการที่จะไปเที่ยวที่จังหวัดภูเก็ตกับครอบครัว จึงต้องการข้อมูลเรื่องโรงแรมที่พักภายในจังหวัดภูเก็ต ซึ่งนายซี ต้องการโรงแรมที่ห้องพักสะอาด เงียบสงบเหมาะแก่การพักผ่อน มีวิวทิวทัศน์ที่สวยงาม และมีลานจอดรถให้บริการอย่างเพียงพอ จึงเข้าไปดูข้อความแสดงความคิดเห็นภายในเว็บไซต์ <http://www.agoda.co.th> ดังแสดงในรูปที่ 3.1

โรงแรมในภูเก็ต | [+ เริ่มการค้นหาใหม่](#) | [รับแต้มสะสมโกด้ารีเวิร์ด - จองเลย!](#) | [รับประกันราคาดีที่สุด](#)

เรียงลำดับตาม	สกุลเงิน	เลือกทำเล	ชื่อโรงแรม (ภาษาอังกฤษเท่านั้น)
ความนิยม	บาท (THB)	- ทุกพื้นที่ -	<input type="text"/> ค้นหา

การโหลดเสร็จสมบูรณ์ < ก่อนหน้า ถัดไป >

153 โรงแรม ใน ภูเก็ต ซึ่งตรงกับค้นหา (จาก 153) สำหรับ 31 ธ.ค. 2553 - 01 ม.ค. 2554



เดอะ ชอร์ แอท กะตะธานี รีสอร์ท (The Shore at Katathani Resort) ★★★★★

18 Kata Noi Road, Muang, ภูเก็ต 83100

ข้อเสนอนี้มีเวลาจำกัด - ราคารวมส่วนลด 25% แล้ว!

แผนที่ สถานที่ตั้ง: กะตะ

โรงแรมยอดเยี่ยมระดับ 5-ดาว แห่งนี้สิ่งอำนวยความสะดวกชั้นยอดมากมายไว้คอยบริการแก่ลูกค้าภายในห้องพักทั้ง 48 ห้อง ภายในห้องพักทุกห้องมี เครื่องปรับอากาศ, เสื้อคลุมอาบน้ำ,....


จองโรงแรมนี้และรับ THB 1,498 สำหรับใช้เป็นส่วนลดในการจองครั้งหน้า [ข้อมูลเพิ่มเติม](#)

1 รีวิวโรงแรม

“ มีความเป็นส่วนตัวในวิลล่าแต่ละหลังค่ะ... [พิมพ์เต็ม](#) ”

เลือก

ยืนยันได้ทันที!



กะตะธานี ภูเก็ต บีช รีสอร์ท (Katathani Phuket Beach Resort) ★★★★★

14

จองด่วน - ราคารวมส่วนลด 23% แล้ว!

แผนที่ สถานที่ตั้ง: กะตะ

โรงแรมระดับ 5-ดาว แห่งนี้ให้บริการห้องพัก 480 ห้อง พร้อมสิ่งอำนวยความสะดวกชั้นยอด ภายในแต่ละห้องมี ห้องปลอดบุหรี่, เครื่องปรับอากาศ, เสื้อคลุมอาบน้ำ, โต๊ะเขียนหนังสือ,....

จองโรงแรมนี้และรับ THB 471 สำหรับใช้เป็นส่วนลดในการจองครั้งหน้า [ข้อมูลเพิ่มเติม](#)

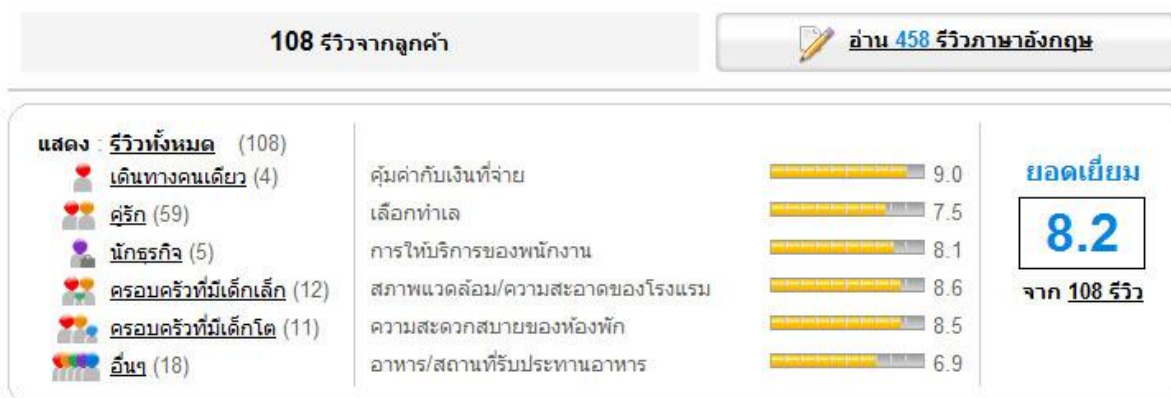
เลือก

ยืนยันได้ทันที!

รูปที่ 3.1 หน้าเว็บไซต์โรงแรมในจังหวัดภูเก็ตจาก <http://www.agoda.co.th>

ภายในเว็บไซต์มีข้อมูลโรงแรมในจังหวัดภูเก็ตมานำเสนอมากมาย และแยกข้อความแสดงความคิดเห็นตามโรงแรมต่างๆ และแต่ละโรงแรมก็มีข้อความแสดงความคิดเห็นเป็นจำนวนมาก ดังแสดงในรูปที่ 3.2 ซึ่งเป็นแถบข้อความแสดงจำนวนข้อความแสดงความคิดเห็นของแต่ละโรงแรมในเว็บไซต์ <http://www.agoda.co.th>

ความคิดเห็นล่าสุดจากลูกค้า



รูปที่ 3.2 จำนวนข้อความแสดงความคิดเห็นของแต่ละโรงแรมในเว็บไซต์ agoda.co.th

สงบเงียบมีความเป็นส่วนตัวสูง

Sompradtana J. ไทย คู่รัก 15 ธันวาคม 2553

3.7

เป็นส่วนตัวดีมาก ไม่วุ่นวาย

ห้องพักธรรมดาหลายๆการบริการไม่สมกับเป็นโรงแรมระดับสี่ดาว จะมีรถรับส่งเฉพาะตอนเช็คอินกับ เช็คเอาท์เท่านั้นนอกนั้นต้องตะกายลงมาเอง ถ้าพักอยู่ห้องบนๆวิวจะสวยแต่เห็นน้อยมากกว่าจะ ถึงลิฟท์ โดยรวมๆแล้วคิดว่าห้องพักและการบริการไม่สมกับราคาเท่าไร

- คำแนะนำเกี่ยวกับโรงแรม : ไม่เหมาะกับครอบครัวที่มีเด็กเล็กหรือมีผู้สูงอายุ

โรงแรมที่ดีและให้บริการด้วยรอยยิ้ม

Suwichan K. ไทย คู่รัก 7 ธันวาคม 2553

8.3

วิวสวย, สงบ, บรรยากาศดี

ที่จอดรถค่อนข้างน้อย ตอนเข้าไปจอดใหม่ๆ งง ว่าตรงไหนเป็นของโรงแรม เพราะมี 2 ระดับ แต่จำรูปได้ก็เลยจอดข้างล่าง ไม่ต้องขึ้นไปข้างบนเพราะไม่มีที่จอด ทางก็ค่อนข้างชัน ส่วนทางไปห้องพักค่อนข้างชันขึ้นเพราะพักอยู่ด้านบน ต้องต่อลิฟท์ 2 ตัว และเดินต่ออีก 2 ช่วง ค่อนข้างเหนื่อยน่าจะไม่เหมาะสำหรับผู้สูงอายุ แต่ดูเหมือนโรงแรมจะมีรถกอล์ฟ น่าจะมีบริการรับรถส่ง แต่ชอบวิวเพราะมันสูงและสวย วิวน้ำค่อนข้างเยอะ ฉะนั้น ดี แต่ก็กลัวๆ ว่าจะไปเล่นน้ำห้องของคนอื่นเค้าเพราะว่ามีสระที่อยู่ด้านหน้าของห้องบางห้องเลย (ประมาณสระส่วนตัว) แต่รู้สึกทางโรงแรมจะมีสระใหญ่อยู่ สำหรับห้องพักค่อนข้างใหญ่ดี สะอาด อุปกรณ์ทุกอย่างดีหมดเลย โดยเฉพาะอินเตอร์เน็ตไว สาย ฟรีด้วย ค่อนข้างชอบ เพราะถ้าต้องทำงานโดยใช้เน็ตจะสะดวกมากมาย โดยรวมแล้วทุกอย่างค่อนข้างดี คิดว่าถ้าได้ไปภูเก็ตอีกน่าจะต้องไปนอนอีกสักครั้ง

- คำแนะนำเกี่ยวกับแหล่งท่องเที่ยว : ขับรถไม่ไกล เดินทางไปจุดชมวิว 3 อ่าวใกล้ แหลมพรหมเทพก็ไม่ไกล, ของกินสามารถหากินได้บริเวณรอบๆ เพราะอยู่ในซอยที่ค่อนข้างคับคั่ง

บนเนินเขา ชมวิว ได้บรรยากาศ หลากรส

Banyen W. ไทย ครอบครัวที่มีเด็กเล็ก 27 ตุลาคม 2553

8.7

การออกแบบสวยยอด

ห้องพักสะอาด พนักงานเป็นกันเอง ไม่แบ่งชั้น ว่าคนไทย/ต่างชาติ ดีมาก ๆ ค่ะ

- คำแนะนำเกี่ยวกับแหล่งท่องเที่ยว : ขับรถชมหาด ต่าง ๆ ดูพระอาทิตย์ตกที่แหลมพรหมเทพ วิวสวยเกิดแฟนตาซี สวยยอด

รูปที่ 3.3 ข้อความแสดงความคิดเห็นของลูกค้าในเว็บไซต์ agoda.co.th

ถึงแม้ว่าในแต่ละข้อความแสดงความคิดเห็นภายในเว็บไซต์ดังในรูปที่ 3.3 จะมีการให้คะแนนโรงแรมนั้นๆ ในมุมมองต่างๆ 6 มุมมองได้แก่ คุ้มค่ากับเงินที่จ่าย, เลือกทำเล, การให้บริการของพนักงาน, สภาพแวดล้อม/ความสะอาดของโรงแรม, ความสะอาดสบายของห้องพัก, อาหาร/สถานที่รับประทานอาหาร

อาหาร แต่ถ้าหากนายซี ต้องการความคิดเห็นที่เป็นรายละเอียดปลีกย่อยจากมุมมองทั้ง 6 มุมมอง หรือ รายละเอียดอื่นที่ไม่เกี่ยวข้องกับมุมมองที่ทางเว็บไซต์กำหนดให้ นายซีก็ต้องอ่านข้อความแสดงความคิดเห็นทั้งหมด เพื่อให้ได้รายละเอียดที่ตรงต่อความต้องการ อาจใช้เวลาในการอ่านข้อความแสดงความคิดเห็นและนำข้อมูลที่ได้มาสรุปเพื่อตัดสินใจในการเลือกโรงแรม เช่น นายซีต้องการโรงแรมที่มีอาหารอร่อย จึงอยากได้ข้อมูลสรุปเรื่องอาหารจากทุกๆ โรงแรม ระบบสามารถวิเคราะห์กรองและสกัดข้อความที่เกี่ยวข้องกับอาหารเท่านั้น เช่น "อาหารเช้าอร่อยมีให้เลือกหลายอย่าง" เป็นความคิดเห็นในเชิงบวก "อาหารเย็นซีด ไม่ได้เรื่อง" เป็นความคิดเห็นในเชิงลบ

ระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรม สามารถสรุปข้อมูลความคิดเห็นเกี่ยวกับการเข้าพักของลูกค้าแต่ละโรงแรมให้นายซีได้ ช่วยให้ประหยัดเวลาในการหาข้อมูลและช่วยให้นายซีสามารถตัดสินใจได้ง่ายขึ้น

3.2 เทคนิคและเทคโนโลยี

ระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรมมีเทคนิคและเทคโนโลยีที่เกี่ยวข้องดังต่อไปนี้

- 1) นิพจน์ปรกติ (Regular Expression)
- 2) การวิเคราะห์ข้อความแสดงความคิดเห็น (Opinion Mining)
- 3) สัญกรณ์วัตถุจาวาสคริปต์ (JavaScript Object Notation - JSON)
- 4) เทคนิคจินตทัศน์ (Information Visualization)

1) นิพจน์ปรกติ (Regular Expression)

การประมวลผลภาษาธรรมชาติภาษาไทย มีความจำเป็นที่จะต้องตัดประโยคภาษาไทยออกเป็นคำๆ แล้วจึงนำไปประมวลผลต่อไป แต่เนื่องจากธรรมชาติของภาษาไทย การแบ่งประโยคหรือแบ่งคำไม่มีรูปแบบที่ตายตัว มักเป็นข้อความที่ยาวต่อกัน มีการแบ่งวรรคตอนเพื่อให้ความหมายเป็นไปอย่างถูกต้อง แต่ก็ขึ้นกับทักษะของผู้เขียนเป็นสำคัญ หากแบ่งวรรคตอนผิด ก็ย่อมส่งผลกระทบต่อการแบ่งประโยคหรือการตัดคำ รวมไปถึงการประมวลผลภาษาธรรมชาติ

ปัจจุบันกลไกการตัดคำภาษาไทย แบ่งออกเป็น 2 วิธีใหญ่ๆ คือการเปรียบเทียบคำจากพจนานุกรมภาษาไทย และการให้ระบบวิเคราะห์ไวยากรณ์ที่เป็นไปได้ในภาษาไทย ซึ่งอย่างหลังเป็นที่นิยมเนื่องจากภาษาไทยมีคำใหม่ๆ เกิดขึ้นมามากมาย จึงเป็นอีกวิธีหนึ่งที่ช่วยลดต้นทุนการเพิ่มเติมคำไทยใหม่ๆ ในคลังคำศัพท์[7]

ระบบวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม ใช้โปรแกรมการตัดคำ LexToPlus ซึ่งพัฒนาโดยหน่วยปฏิบัติการวิจัยวิทยาการมนุษยภาษา ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ในการตัดคำสำหรับข้อความแสดงความคิดเห็นจากเว็บไซต์ <http://www.agoda.co.th>

ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรมที่นำมา นั้น เป็นข้อความแสดงความคิดเห็นที่ลูกค้าของโรงแรมที่จองห้องพักผ่านเว็บไซต์ <http://www.agoda.co.th> เป็นผู้เขียน จึงมีลักษณะที่เป็นภาษาธรรมชาติ ก่อนนำไปประมวลผลจึงต้องมีการแบ่งประโยคและตัดคำเสียก่อน ผู้พัฒนาจึงนำนิพจน์ปรกติมาประยุกต์ใช้ในการแบ่งประโยคข้อความแสดงความคิดเห็นจากข้อความแสดงความคิดเห็นทั้งหมด

นิพจน์ปรกติ ใช้สำหรับการอธิบายรูปแบบของข้อความ(String) มักใช้ในการกำหนดรูปแบบเพื่อค้นหาข้อความหรือตัวอักษรว่ามีอยู่ในข้อความที่กำหนดหรือไม่ รูปแบบไวยากรณ์ของนิพจน์ปรกติที่ควรรู้จักมีดังนี้ [9]

1. การกำหนดว่า ต้องเป็นตัวแรกของข้อความ

เมื่อต้องการตรวจสอบว่า ข้อความที่กำลังตรวจสอบ ขึ้นต้นด้วยคำที่ต้องการหรือไม่ จะใช้เครื่องหมาย ^ เพื่อค้นหาว่า ข้อความที่จะค้นหานั้นขึ้นต้นด้วยคำที่กำหนดหรือไม่

ตัวอย่าง นิพจน์ปรกติ คือ "^ส"

หากนำไปค้นหาในข้อความ "สวัสดีค่ะ" จะได้ค่าความจริงเป็นจริง

หากนำไปค้นหาในข้อความ "คุณครู สวัสดีค่ะ" จะได้ค่าความจริงเป็นเท็จ

2. การกำหนดว่า ต้องเป็นตัวสุดท้ายของข้อความ

เมื่อต้องการตรวจสอบว่า ข้อความที่กำลังจะตรวจสอบ ลงท้ายด้วยคำที่ต้องการหรือไม่ จะใช้เครื่องหมาย \$ เพื่อค้นหาว่า ข้อความที่จะค้นหานั้นลงท้ายด้วยคำที่กำหนดหรือไม่

ตัวอย่าง นิพจน์ปรกติ คือ "บ\$"

หากนำไปค้นหาในข้อความ "สวัสดีครับ" จะได้ค่าความจริงเป็นจริง

หากนำไปค้นหาในข้อความ "สวัสดีค่ะ" จะได้ค่าความจริงเป็นเท็จ

3. การค้นหาตัวอักษร:

เมื่อต้องการตรวจสอบว่า จะต้องมีตัวอักษรจะเป็นตัวเลขหรือตัวอักษรก็ได้ ซึ่งตัวอักษระนั้นต้องไม่ใช่ช่องว่าง จะใช้เครื่องหมาย . เพื่อค้นหาว่า ข้อความที่จะค้นหานั้นมีตัวอักษรหรือตัวเลขประกอบอยู่หรือไม่

ตัวอย่าง นิพจน์ปรกติ คือ "^."

หากนำไปค้นหาในข้อความ "1 2 3 สวัสดีค่ะ" จะได้ค่าความจริงเป็นจริง

หากนำไปค้นหาในข้อความ "สวัสดีครับ" จะได้ค่าความจริงเป็นเท็จ

4. การกำหนดว่า อาจจะมีตัวอักษรที่กำหนดหรือไม่ก็ได้

ใช้เครื่องหมาย * เพื่อระบุว่าในข้อความที่ต้องการค้นหาอาจจะมีหรือไม่มีข้อความที่กำหนดก็ได้ ถ้ามี ก็อาจมีได้หลายตัว

ตัวอย่าง นิพจน์ปรกติ คือ "ร*"

หากนำไปค้นหาในข้อความ "ข้อความแสดงความคิดเห็น" จะได้ค่าความจริงเป็นจริง

หากนำไปค้นหาในข้อความ "ระบบวิเคราะห์ข้อความ" จะได้ค่าความจริงเป็นจริง

หากนำไปค้นหาในข้อความ "คณะกรรมการ" จะได้ค่าความจริงเป็นจริง

ใช้เครื่องหมาย ? เพื่อระบุว่าในข้อความที่ต้องการค้นหาอาจจะมีหรือไม่มีข้อความที่กำหนดก็ได้ ถ้ามี ก็อาจมีได้เพียงตัวเดียว

ตัวอย่าง นิพจน์ปรกติ คือ "ร?"

หากนำไปค้นหาในข้อความ "ข้อความแสดงความคิดเห็น" จะได้ค่าความจริงเป็นจริง

หากนำไปค้นหาในข้อความ "ระบบวิเคราะห์ข้อความ" จะได้ค่าความจริงเป็นจริง

หากนำไปค้นหาในข้อความ "บ้านจัดสรร" จะได้ค่าความจริงเป็นเท็จ

5. การกำหนดให้มีตัวอักษรที่กำหนดอย่างน้อย 1 ตัว

ใช้เครื่องหมาย + เพื่อระบุว่าในข้อความที่ต้องการค้นหาจะต้องมีตัวอักษรนั้นอย่างน้อย 1 ตัว

ตัวอย่าง นิพจน์ปกติ คือ "ร?"

หากนำไปค้นหาในข้อความ "ระบบวิเคราะห์ข้อความ" จะได้ค่าความจริงเป็นจริง

หากนำไปค้นหาในข้อความ "บ้านจัดสรร" จะได้ค่าความจริงเป็นจริง

หากนำไปค้นหาในข้อความ "ข้อความแสดงความคิดเห็น" จะได้ค่าความจริงเป็นเท็จ

6. การตรวจสอบช่องว่าง

ใช้ `[:space:]` หรืออาจจะใช้ " " แทนก็ได้ แต่การใช้ `:space:` จะหมายความรวมถึงเครื่องหมายอื่น ๆ เช่น การย่อหน้า เครื่องหมายขึ้นบรรทัดใหม่ รวมทั้งการเว้นวรรคด้วย

ตัวอย่าง นิพจน์ปกติ คือ `[:space:]`

หากนำไปค้นหาในข้อความ "ระบบ วิเคราะห์ข้อความ" จะได้ค่าความจริงเป็นจริง

หากนำไปค้นหาในข้อความ "ข้อความแสดงความคิดเห็น" จะได้ค่าความจริงเป็นเท็จ

7. การตรวจสอบเป็นช่วง

ใช้เครื่องหมาย `[และ]` เพื่อคร่อมช่วงที่ต้องการ เช่น ต้องการระบุว่า ต้องเป็นระหว่างเลข 0 – 9 เท่านั้น เราจะเขียนว่า `[0-9]`

ตัวอย่าง นิพจน์ปกติ คือ `[ก-๙]` (กำหนดให้มีเฉพาะตัวอักษรภาษาไทยเท่านั้น มีตัวเลขที่เป็นภาษาไทยได้ แต่ห้ามมีตัวอักษรภาษาอังกฤษ)

หากนำไปค้นหาในข้อความ "ข้อความแสดงความคิดเห็น" จะได้ค่าความจริงเป็นจริง

หากนำไปค้นหาในข้อความ " Alphabet" จะได้ค่าความจริงเป็นเท็จ

ในกรณีของตัวอักษรภาษาไทย ถ้าจะกำหนดไม่ให้มีตัวเลข ต้องกำหนดค่า ระหว่าง ตัว ก ไก่ และ ไม้ตรี ถ้ากำหนดตั้งแต่ ตัว ก-ฮ จะทำให้มีสระไม่ได้ ซึ่งก็จะทำให้ไม่สามารถสะกดตัวเป็นคำได้อย่างสมบูรณ์

8. การจัดกลุ่มคำ หรือ ข้อความ

ใช้เครื่องหมายวงเล็บ () คร่อม

ตัวอย่าง นิพจน์ปกติ คือ `^(การ)`

หากนำไปค้นหาในข้อความ "การแสดงความคิดเห็น" จะได้ค่าความจริงเป็นจริง

หากนำไปค้นหาในข้อความ "เสนอความคิดเห็น" จะได้ค่าความจริงเป็นเท็จ

9. การตรวจสอบเครื่องหมาย หรือ สัญลักษณ์พิเศษ

เนื่องจากในนิพจน์ปกติใช้เครื่องหมายในไวยากรณ์ ดังนั้นหากต้องการค้นหาเครื่องหมายดังกล่าวในข้อความที่กำหนด จึงจำเป็นต้องบอกให้โปรแกรมรู้ว่า เครื่องหมายที่ใช้ไม่ใช่เป็นส่วนหนึ่งของไวยากรณ์ เรียกว่าเป็นการ escaping คือการใช้เครื่องหมาย \ ใส่ไว้หน้าตัวอักษรพิเศษหรือสัญลักษณ์นั้นๆ เช่น * เพื่อระบุว่า ให้ตรวจหาเครื่องหมาย * ในข้อความที่ต้องการค้นหา

ระบบวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม ใช้นิพจน์ปกติในการกำหนดรูปแบบประโยคข้อความแสดงความคิดเห็น เมื่อข้อความแสดงความคิดเห็นถูกตัดคำออกมาด้วยโปรแกรม LexToPlus แล้วจึงนำคำที่ถูกตัด มากำกับตามประเภทคำดังนี้

FEA – คือคำระบุคุณลักษณะของโรงแรม ซึ่งแบ่งเป็น 8 ประเภท ได้แก่ การบริการ, อาหาร, สภาพโรงแรม, สถานที่ตั้ง, ห้องพัก, สิ่งอำนวยความสะดวก, ภาพรวมอื่นๆ

POL – คือคำระบุข้อความความคิดเห็น ซึ่งจะทำให้ข้อความแสดงความคิดเห็นสามารถระบุว่ามีความคิดเห็นในเชิงบวก เชิงลบ หรือเป็นกลางได้

DEG – คือคำระบุปริมาณ ซึ่งมักกำกับอยู่กับคำระบุข้อความความคิดเห็น

NEG – คือคำที่มีความหมายในเชิงลบ จากการศึกษาระบบวิเคราะห์ข้อความแสดงความคิดเห็น คำที่มีความหมายในเชิงลบมักใช้หน้าคำระบุข้อความความคิดเห็น และทำให้คำระบุข้อความความคิดเห็นมีความหมายตรงกันข้าม

STO – คือคำหยุด ซึ่งมักจะไม่มี ความหมายในประโยค ในที่นี้หมายถึงคำว่า “การ” หรือ “ความ” ซึ่งเมื่อนำมานำหน้าคำกริยาแล้วสามารถกลายเป็นคำนามในประโยคได้ จากการศึกษาระบบวิเคราะห์ข้อความแสดงความคิดเห็น คำระบุข้อความความคิดเห็นที่ถูกลำหน้าด้วยคำว่า “การ” หรือ “ความ” นี้สามารถนำมาเป็นคำระบุคุณลักษณะของโรงแรมได้

HAV – คือคำว่า “มี” ซึ่งจากการศึกษาระบบวิเคราะห์ข้อความแสดงความคิดเห็น คำว่า “มี” ในบางกรณีเมื่อนำหน้าคำระบุคุณลักษณะแล้วสามารถกลายเป็นคำระบุข้อความความคิดเห็นได้ เช่น “พนักงานมีความกระตือรือร้น” หรืออาจมีความหมายเป็นคำระบุคุณลักษณะเช่นเดิมได้ เช่น “มีการตกแต่งห้องพักที่สวยงาม”

AUX – คือคำกริยาช่วย มักใช้กับภาคแสดงของประโยค เช่น พนักงานควรจะช่วยเหลือแขกที่มาพัก คำว่า “ควรจะ” จัดเป็นคำในประเภทของ AUX

PRE – คือคำที่เชื่อมคำสองคำเข้าด้วยกัน เช่นคำว่า “และ”, “กับ”, “ของ” เป็นต้น

CON – คือคำเชื่อมประโยค หรือคำสันธาน ระบบวิเคราะห์ข้อความแสดงความคิดเห็นใช้คำเชื่อมประโยคในการแบ่งข้อความแสดงความคิดเห็นประโยคย่อยๆออกจากข้อความแสดงความคิดเห็นที่ได้จากเว็บไซต์ <http://www.agoda.co.th>

PAR – คือคำที่ใช้ในการลงท้ายประโยค ยกตัวอย่างเช่นคำว่า “เลย”, “คะ”, “ด้วย” เป็นต้น

OTH – คือคำที่ระบบไม่สามารถระบุประเภทได้

ระบบวิเคราะห์ข้อความแสดงความคิดเห็นใช้คำทั้ง 11 ประเภทนี้ แทนคำแต่ละคำของข้อความแสดงความคิดเห็นที่นำมาวิเคราะห์ จากนั้นจึงนำข้อความที่ถูกแทนด้วยประเภทของคำไปค้นหาเปรียบเทียบกับรูปแบบข้อความแสดงความคิดเห็นที่ระบบวิเคราะห์ข้อความแสดงความคิดเห็นกำหนดไว้

2) การวิเคราะห์ข้อความแสดงความคิดเห็น (Opinion Mining)

ในปัจจุบันเว็บไซต์โดยทั่วไปมีลักษณะที่ส่งเสริมการแบ่งปันข้อมูล การพัฒนาในด้านแนวความคิดและการออกแบบ รวมถึงการร่วมสร้างข้อมูลบนโลกของอินเทอร์เน็ต แทนที่แนวคิดของเว็บไซต์ในแบบเก่าที่เป็นลักษณะการให้บริการอ่านเพียงอย่างเดียว ดังนั้นข้อมูลบนเว็บไซต์จึงแบ่งได้ออกเป็น 2 รูปแบบคือ

1. ข้อเท็จจริง เป็นข้อมูลจริงที่สามารถพิสูจน์ได้ เช่น โลกเป็นดาวเคราะห์ที่อยู่ห่างจากดวงอาทิตย์เป็นลำดับที่สาม

2. ข้อคิดเห็น คือการแสดงทัศนคติที่ผู้เขียนมีต่อสิ่งๆหนึ่งหรือหัวข้อๆหนึ่ง โดยแบ่งออกเป็น

- ความคิดเห็นแบบตรง (Direct opinion) สามารถบอกทัศนคติของผู้เขียนหรือผู้พูดได้อย่างชัดเจน เช่น “รูปที่ถ่ายจากกล้องตัวนี้ คุณภาพดี”

- ความคิดเห็นเชิงเปรียบเทียบ (Comparisons) คือการให้ความเห็นโดยนำสิ่งที่ต้องการแสดงความคิดเห็นไปเปรียบเทียบกับอีกสิ่งหนึ่ง เช่น “กล้อง A แพงกว่ากล้อง B” [4]

องค์ประกอบของการแสดงความคิดเห็น [2] ประกอบด้วย 3 ส่วนดังปรากฏในรูปที่ 3.4 คือ

1. ผู้แสดงความคิดเห็น (Opinion Holder)

2. วัตถุหรือหัวข้อที่นำมาแสดงความคิดเห็น (Object)

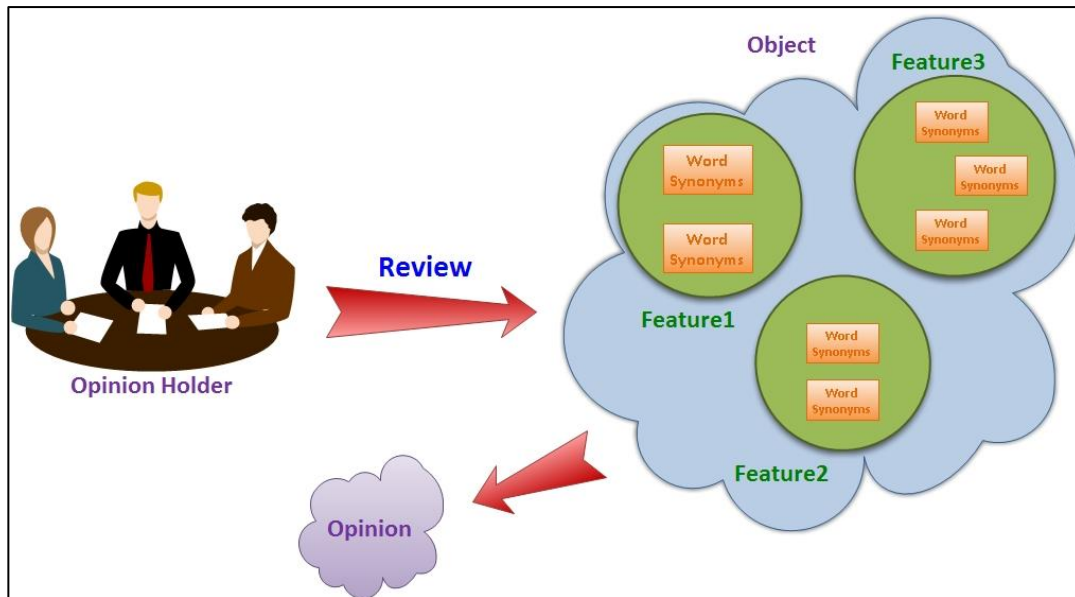
3. ความคิดเห็นหรือทัศนคติของผู้แสดงความคิดเห็นที่มีต่อหัวข้อที่นำมาแสดงความคิดเห็น (Opinion) ซึ่งแต่ละหัวข้อที่ถูกลำหน้ามาแสดงความคิดเห็นจะประกอบไปด้วยคุณลักษณะต่างๆ (feature) สามารถเขียนในรูปแบบของเซตข้อมูลได้ดังนี้

Object = {F₁, F₂, F₃, ..., F_n}

โดย F_n คือ คุณลักษณะที่วัตถุหรือหัวข้อที่นำมาแสดงความคิดเห็นมี ซึ่งสามารถแทนได้ด้วยเซตของข้อมูล

$$F_n = \{W_1, W_2, W_3, \dots, W_n\}$$

โดย W_n คือ คำที่มีความหมายเหมือนกัน(word synonyms) มีความหมายที่สามารถแทนที่กับคุณลักษณะนั้นๆได้



รูปที่ 3.4 องค์ประกอบของการแสดงความคิดเห็น

เมื่อผู้ให้ความคิดเห็นแสดงความคิดเห็นเกี่ยวกับวัตถุหรือหัวข้อบนคุณลักษณะที่เป็นส่วนหนึ่งของวัตถุหรือหัวข้อนั้นๆ ก็จะถูกนำไปตรวจสอบกับคำที่มีความหมายเหมือนในแต่ละคุณลักษณะแล้วนำมาตัดสินว่าเป็นการแสดงความคิดเห็นในข้อใด ซึ่งอาจจะเป็นข้อบวก(positive) ข้อลบ(negative) หรือเป็นกลาง(neutral) ดังลักษณะของข้อความแสดงความคิดเห็นในรูปที่ 3.5 [1,3]



รูปที่ 3.5 ตัวอย่างรูปแบบข้อความแสดงความคิดเห็น

ซึ่งระบบวิเคราะห์ข้อความแสดงความคิดเห็น คือ กระบวนการอัตโนมัติเพื่อใช้ตรวจสอบทัศนคติของผู้พูดหรือผู้เขียนในหัวข้อเรื่องใดเรื่องหนึ่ง โดยอาศัยการสนทนาคอมพิวเตอร์ให้พิจารณาอารมณ์ความรู้สึกด้วยการประมวลผลด้วยภาษาธรรมชาติ (Natural Language Processing, NLP) แบ่งการทำงานออกเป็น 3 ระดับคือ

1. ระดับเอกสาร (Document Level) เป็นการวิเคราะห์ข้อความแสดงความคิดเห็นในแบบหยาบ เนื่องจากเป็นการนำข้อความแสดงความคิดเห็นทั้งหมดจากเอกสาร มาสรุปแยกข้อความความคิดเห็นเป็นขั้วบวก ขั้วลบ หรือเป็นกลาง
2. ระดับประโยค (Sentence Level) เป็นการวิเคราะห์ข้อความแสดงความคิดเห็น โดยแยกข้อความที่เป็นข้อความแสดงความคิดเห็นออกมาจากข้อความที่เป็นข้อเท็จจริงในระดับที่เป็นประโยค แล้วนำมาแยกข้อความความคิดเห็นเป็นขั้วบวก ขั้วลบ หรือเป็นกลาง
3. ระดับคุณลักษณะ (Feature Level) เป็นการวิเคราะห์ข้อความแสดงความคิดเห็น โดยแยกลักษณะที่สนใจของวัตถุหรือหัวข้อที่ถูกแสดงความคิดเห็นออกมาก่อน แล้วจึงนำมาแบ่งข้อความความคิดเห็นเป็นขั้วบวก ขั้วลบ หรือเป็นกลาง และนำมาจัดกลุ่มเข้ากับคำที่มีความหมายเหมือนในแต่ละคุณลักษณะ

ซึ่งระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรม จะนำข้อความแสดงความคิดเห็นของลูกค้ามาวิเคราะห์ข้อความแสดงความคิดเห็นในระดับคุณลักษณะ แล้วนำข้อมูลที่ได้มาประมวลผลเพื่อให้ผู้ใช้เข้าใจได้ง่ายขึ้น

3) สัญกรณ์วัตถุจาวาสคริปต์ (JavaScript Object Notation - JSON)

เป็นรูปแบบที่ใช้สำหรับแลกเปลี่ยนข้อมูลทางคอมพิวเตอร์ อยู่ในรูปข้อความธรรมดา(Plain text) ที่ทั้งมนุษย์และโปรแกรมคอมพิวเตอร์สามารถอ่านเข้าใจได้ ใช้รูปแบบของภาษาจาวาสคริปต์ แต่ไม่ถูกมองว่าเป็นภาษาโปรแกรม ปัจจุบันเป็นที่นิยมใช้ในเว็บแอปพลิเคชันโดยเฉพาะ Asynchronous JavaScript And XML(AJAX) ที่ใช้สัญกรณ์วัตถุจาวาสคริปต์ เป็นทางเลือกในการส่งข้อมูล เพิ่มจาก Extensible Markup Language(XML) ที่เป็นที่นิยมใช้กันอยู่แต่เดิม สาเหตุที่สัญกรณ์วัตถุจาวาสคริปต์ เริ่มได้รับความนิยมเป็นเพราะกระชับและเข้าใจง่ายกว่า XML ซึ่งตัวอย่างการเปรียบเทียบเป็นไปตามรูปที่ 3.6 และรูปที่ 3.7

สัญกรณ์วัตถุจาวาสคริปต์ใช้ [] แทนอาร์เรย์ และ { } แทน Hash หรือ Associative array(ประเภทข้อมูลอย่างย่อ ที่สามารถเข้าถึงข้อมูลผ่านข้อมูลอีกตัว โดยผ่านคีย์) [8]

ระบบวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม ใช้รูปแบบการแลกเปลี่ยนข้อมูลแบบสัญกรณ์วัตถุจาวาสคริปต์ในการแปลงข้อความแสดงความคิดเห็นที่วิเคราะห์แล้วให้เป็นรูปแบบการแสดงผลแบบกราฟิก

ตัวอย่างรูปแบบของ XML

```
<profile>
  <person>
    <firstname>Scalet</firstname>
    <lastname>Ohara</lastname>
  </person>
  <person>
    <firstname>Mary</firstname>
    <lastname>Jane</lastname>
  </person>
</profile>
```

รูปที่ 3.6 ตัวอย่างรูปแบบ xml

ตัวอย่างรูปแบบของ JSON เมื่อแปลงจาก XML ด้านบน

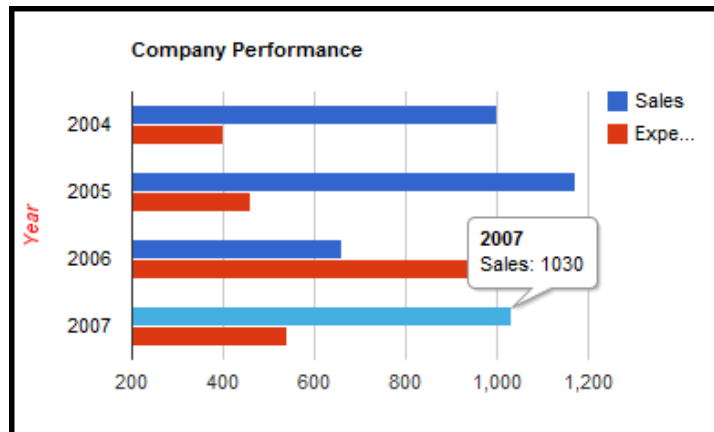
```
{
  "person" :
    [
      { "firstname" : "Scalet", "lastname" : "Ohara" },
      { "firstname" : "Mary", "lastname" : "Jane" }
    ]
}
```

รูปที่ 3.7 ตัวอย่างรูปแบบ JSON

4) เทคนิคจินตทัศน์ (Information Visualization)

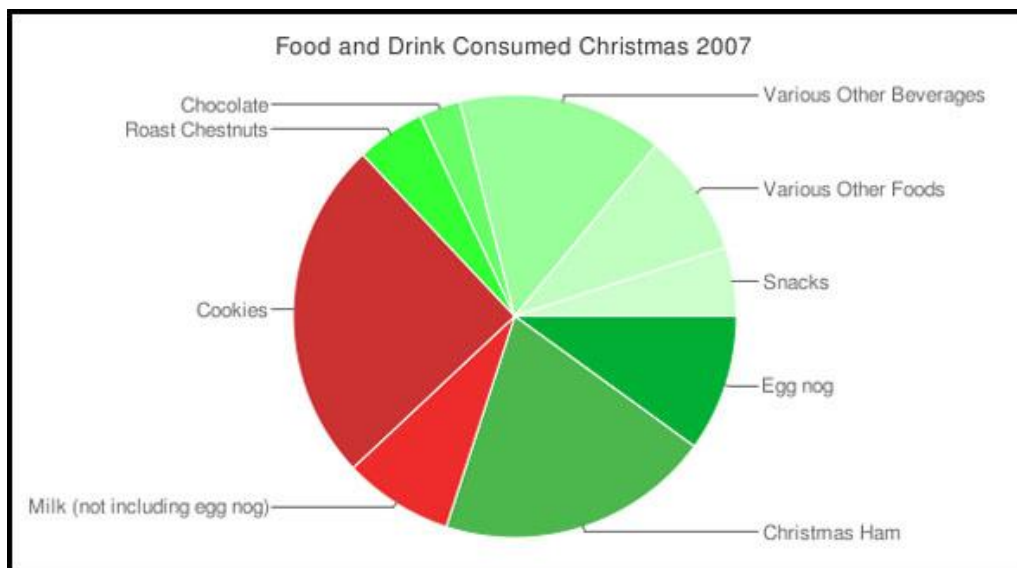
เทคนิคจินตทัศน์ คือ การนำเสนอข้อมูลที่เข้าใจยาก เช่นตัวเลข หรือ ตัวหนังสือ ที่มีปริมาณมากมาย มาแสดงให้อยู่ในรูปแบบที่ผู้อ่านเข้าใจได้ง่ายกว่าเดิม เช่น กราฟ แผนภาพแบบต่างๆ โดยเลือกใช้วิธีที่จะนำเสนอให้ผู้อ่านเข้าใจได้ง่ายที่สุด ซึ่งขึ้นกับข้อมูลที่จะนำมานำเสนอ ว่าจะเป็นข้อมูลที่อยู่ในลักษณะใด [5]

การนำเสนอข้อมูลผ่านเทคนิคจินตทัศน์ทำได้หลายรูปแบบ แต่ที่นิยมใช้กันมากนั้นคือการนำเสนอในรูปแบบที่เป็น 2 มิติ เช่น Bar Chart, Pie Chart เป็นต้น



รูปที่ 3.8 เทคนิคจินตทัศน์แบบ Bar Chart 2 มิติ (ภาพจาก Google Visualization)

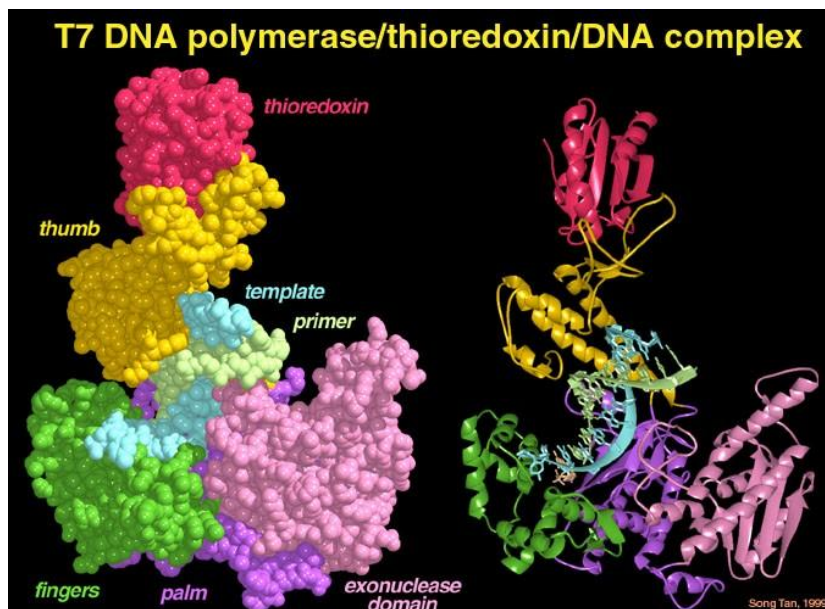
จากรูปที่ 3.8 แสดงถึงสถิติยอดขายและค่าใช้จ่ายของบริษัทซึ่งสามารถทำให้ผู้อ่านเข้าใจได้ง่ายกว่าการนำตัวเลขมาแสดงโดยตรง



รูปที่ 3.9 เทคนิคจินตทัศน์แบบ Pie Chart 2 มิติ (ภาพจาก <http://24ways.org>)

จากรูปที่ 3.9 แสดงถึงปริมาณอาหารและเครื่องดื่มที่ใช้บริโภคในเทศกาลคริสต์มาส ปี 2007 ซึ่งสามารถแสดงให้เห็นอย่างชัดเจนถึงปริมาณที่แตกต่างกันของอาหารและเครื่องดื่มแต่ละประเภท

ในรูปที่ 3.10 เป็นรูปแบบการนำเสนอแบบ 3 มิติ ในตอนต้นถูกนำมาใช้เพื่อความสวยงาม แต่ในแง่ของการใช้ประโยชน์นั้น ยังคงเป็นข้อมูลที่มีเพียง 1 หรือ 2 มิติเท่านั้น แต่ในปัจจุบันรูปแบบการนำเสนอแบบ 3 มิติถูกนำมาใช้ประโยชน์มากขึ้น เช่น การนำเสนอข้อมูลเชิงวิทยาศาสตร์ ในรูปที่ 8 [6]



รูปที่ 3.10 เทคนิคจินตทัศน์แบบ 3 มิติ ของโครงสร้าง DNA (ภาพจาก life.illinois.edu)

ระบบวิเคราะห์ที่ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม ใช้เทคนิคจินตทัศน์แบบ Pie Chart ในการแสดงข้อมูลสรุปข้อความแสดงความคิดเห็นตามคุณลักษณะของแต่ละโรงแรม และใช้เทคนิคจินตทัศน์แบบ Bar Chart ในการแสดงข้อมูลเปรียบเทียบข้อความแสดงความคิดเห็นตามคุณลักษณะของโรงแรมที่เลือก

3.3 เครื่องมือที่ใช้ในการพัฒนา

ภาษาที่ใช้ในการพัฒนา

1. Java ใช้ในการพัฒนาระบบวิเคราะห์ข้อความแสดงความคิดเห็น
2. JSP ใช้ในการเขียนหน้าเว็บไซต์เพื่อรับข้อมูลจากผู้นนำมาประมวลผล
3. HTML ใช้ในการเขียนหน้าเว็บไซต์เพื่อติดต่อกับผู้ใช้
4. Java Script ใช้ในการสร้างการแสดงผลข้อมูลด้วยเทคนิคจินตทัศน์

เครื่องมือที่ใช้ในการพัฒนา

1. Eclipse Java EE IDE for Web Developers (Open Source)
2. AppServ (Open Source)
3. JavaScript InfoVis Toolkit - thejit (Open Source)
4. Apache Tomcat 6.0 (Open source)
5. LexToPlus (Open Source)

ไลบรารีที่ใช้ในการพัฒนา

1. Mysql Connector สำหรับการติดต่อบริการข้อมูล SQL ด้วยภาษา Java
2. Java Excel API (JXL) สำหรับการอ่านไฟล์ Spreadsheet

รายละเอียดโปรแกรมที่จะพัฒนา

- ข้อมูลรับเข้า
 - สถานที่
 - ชื่อโรงแรม
 - มุมมองที่ผู้ใช้สนใจ
- ข้อมูลผลลัพธ์
 - กราฟิกสรุปข้อมูลจากข้อความแสดงความคิดเห็นของสถานที่ หรือโรงแรมนั้นๆ
 - ข้อความแสดงความคิดเห็นที่ถูกลำมาใช้สรุปข้อมูลแยกตามมุมมองที่ผู้ใช้สนใจ
- ฟังก์ชันการทำงาน

ภายในคลาส OpinionPatternAnalysis มีฟังก์ชันการทำงานให้เรียกใช้ดังปรากฏในรูปที่ 3.11

ดังนี้

OpinionPatternAnalysis		
GenerateWordSegment		
input	parameters	GenerateWordSegment
output	parameters	GenerateWordSegmentResponse
GenerateInputPattern		
input	parameters	GenerateInputPattern
output	parameters	GenerateInputPatternResponse
GetDualPattern		
input	parameters	GetDualPattern
output	parameters	GetDualPatternResponse
GenerateOpinionSegment		
input	parameters	GenerateOpinionSegment
output	parameters	GenerateOpinionSegmentResponse
ExtractUnknown		
input	parameters	ExtractUnknown
output	parameters	ExtractUnknownResponse
IdentifyFeature		
input	parameters	IdentifyFeature
output	parameters	IdentifyFeatureResponse
IdentifyPolar		
input	parameters	IdentifyPolar
output	parameters	IdentifyPolarResponse
ExtractOpinion		
input	parameters	ExtractOpinion
output	parameters	ExtractOpinionResponse

รูปที่ 3.11 ฟังก์ชันการทำงานในคลาส OpinionPatternAnalysis

1. ฟังก์ชัน GenerateWordSegment()

ข้อมูลนำเข้า : ข้อความตัวอักษร

ข้อมูลส่งออก : ArrayList ประเภท String ของข้อความที่ถูกตัดคำแล้ว

การทำงาน : รับข้อความเพื่อนำไปตัดคำ

2. ฟังก์ชัน GenerateInputPattern()

ข้อมูลนำเข้า : ArrayList ประเภท String ของข้อความต้องการระบุประเภทคำ

ข้อมูลส่งออก : ArrayList ประเภท String ของข้อความที่ระบุประเภทคำ

การทำงาน : นำข้อความที่ตัดคำแล้วไประบุประเภทของคำตามคลังข้อความของระบบวิเคราะห์ข้อความแสดงความคิดเห็น

3. ฟังก์ชัน GetDualPattern()

ข้อมูลนำเข้า : ข้อความตัวอักษรที่มีคำที่ระบบไม่รู้จัก

ข้อมูลส่งออก : ข้อความตัวอักษรที่ระบุประเภทของคำที่ระบบประมวลผล

การทำงาน : นำข้อความที่มีคำไม่มีในคลังข้อความของระบบวิเคราะห์ข้อความแสดงความคิดเห็นมาประมวลผลเพื่อหาความเป็นไปได้ที่จะเป็นคำระบุคุณลักษณะหรือคำระบุข้อความความคิดเห็น

4. ฟังก์ชัน GenerateOpinionSegment()

ข้อมูลนำเข้า : ArrayList ประเภท String 2 ชุดข้อมูล คือ ชุดที่เป็นรูปแบบประโยคและชุดที่เป็นคำของรูปแบบประโยค

ข้อมูลส่งออก : ข้อความตัวอักษร ในรูปแบบของอาเรย์ 2 มิติ ประกอบด้วยชุดรูปแบบของประโยคและชุดคำของประโยค

การทำงาน : นำข้อความที่มีคำไม่มีในคลังข้อความของระบบวิเคราะห์ข้อความแสดงความคิดเห็นมาประมวลผลเพื่อหาความเป็นไปได้ที่จะเป็นคำระบุคุณลักษณะหรือคำระบุข้อความความคิดเห็น

5. ฟังก์ชัน ExtractUnknown()

ข้อมูลนำเข้า : ข้อความตัวอักษรที่ต้องการระบุประเภทของคำ จากคำที่ยังไม่มีในคลังข้อความของระบบวิเคราะห์ข้อความแสดงความคิดเห็น

ข้อมูลส่งออก : ชุดข้อความตัวอักษรที่ระบุประเภทของคำที่ยังไม่มีในคลังข้อความของระบบวิเคราะห์ข้อความแสดงความคิดเห็น

การทำงาน : นำคำที่ยังไม่มีในคลังข้อความที่ระบบวิเคราะห์ข้อความแสดงความคิดเห็นวิเคราะห์ได้มาวิเคราะห์ว่าเป็นคำระบุคุณลักษณะหรือคำระบุข้อความความคิดเห็น

6. ฟังก์ชัน IdentifyFeature()

ข้อมูลนำเข้า : ข้อความตัวอักษรที่ต้องการระบุความหมายตามประเภทคุณลักษณะของโรงแรม

ข้อมูลส่งออก : ตัวเลขจำนวนเต็มที่ระบุประเภทคุณลักษณะของโรงแรม

การทำงาน : นำข้อความมาประมวลผลเพื่อระบุประเภทคุณลักษณะของโรงแรม

7. ฟังก์ชัน IdentifyPolar()

ข้อมูลนำเข้า : ข้อความตัวอักษรที่ต้องการระบุข้อความคิดเห็น และ ตัวเลขจำนวนเต็มระบุประเภทคุณลักษณะของโรงแรม

ข้อมูลส่งออก : ตัวเลขจำนวนเต็มที่ระบุข้อความคิดเห็น

การทำงาน : นำข้อความมาประมวลผลเพื่อระบุข้อความคิดเห็น

8. ฟังก์ชัน ExtractOpinion ()

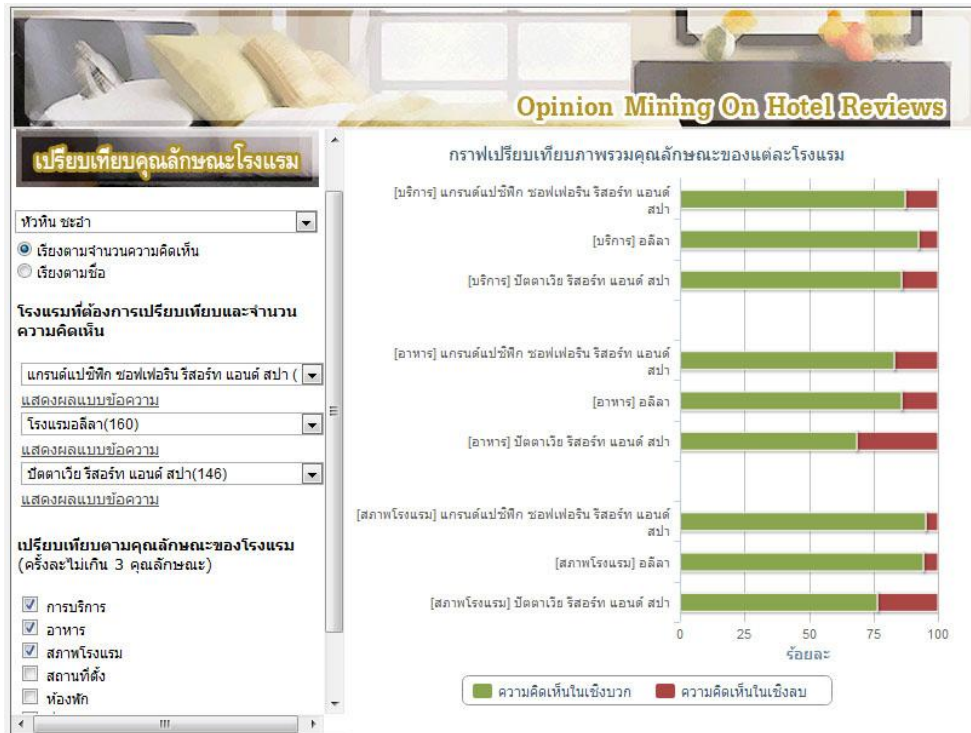
ข้อมูลนำเข้า : ข้อความตัวอักษรที่ต้องการวิเคราะห์ข้อความแสดงความคิดเห็น

ข้อมูลส่งออก : ArrayList ประเภท OpinionResult ซึ่งประกอบด้วย รูปแบบของข้อความแสดงความคิดเห็น, ข้อความแสดงความคิดเห็นที่สกัดได้, คำระบุคุณลักษณะที่สามารถสกัดเพิ่มเติมได้, คำระบุข้อความคิดเห็นที่สามารถสกัดเพิ่มเติมได้, ประเภทคุณลักษณะของข้อความที่วิเคราะห์ และข้อความคิดเห็นของข้อความที่วิเคราะห์

การทำงาน : นำข้อความมาประมวลผลเพื่อแยกข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม ออกตามคุณลักษณะของโรงแรม พร้อมกับระบุข้อความคิดเห็นของข้อความ

เมื่อนำการทำงานของคลาส OpinionPatternAnalysis มาใช้ในการแสดงผล ทำให้มีฟังก์ชันการทำงานดังนี้

1. แสดงข้อมูลทางสถิติสรุปความคิดเห็นจากการโรงแรมที่อยู่ในสถานที่ที่ผู้ใช้สนใจ เปรียบเทียบโดยคุณลักษณะของโรงแรมที่ผู้ใช้เลือก ตามรูปที่ 3.12



รูปที่ 3.12 สรุปความคิดเห็นจากการโรงแรมที่อยู่ในสถานที่ที่ผู้ใช้สนใจ

2. แสดงข้อมูลทางสถิติสรุปความคิดเห็นที่เกี่ยวกับคุณลักษณะต่างๆของโรงแรมที่เลือก ดังเช่นในรูปที่ 3.13



รูปที่ 3.13 สรุปความคิดเห็นที่เกี่ยวกับคุณลักษณะต่างๆของโรงแรมที่เลือก

3. แสดงข้อความแสดงความคิดเห็นที่นำมาสรุปเป็นข้อมูลทางสถิติ ในรูปที่ 3.14

Opinion Mining On Hotel Reviews

คุณลักษณะของโรงแรม

กรุงเทพฯ

เรียงตามจำนวนความคิดเห็น

เรียงตามชื่อ

รายชื่อโรงแรมและจำนวนความคิดเห็น

อะบิลุม เอ็กซ์คลูซีฟ เซอร์วิส อพาร์ทเมนท์(112)

แสดงผลแบบกราฟิก

แสดงผลแบบข้อความ

แสดงคำบอกคุณลักษณะ

แสดงคำระบุข้อความความคิดเห็น

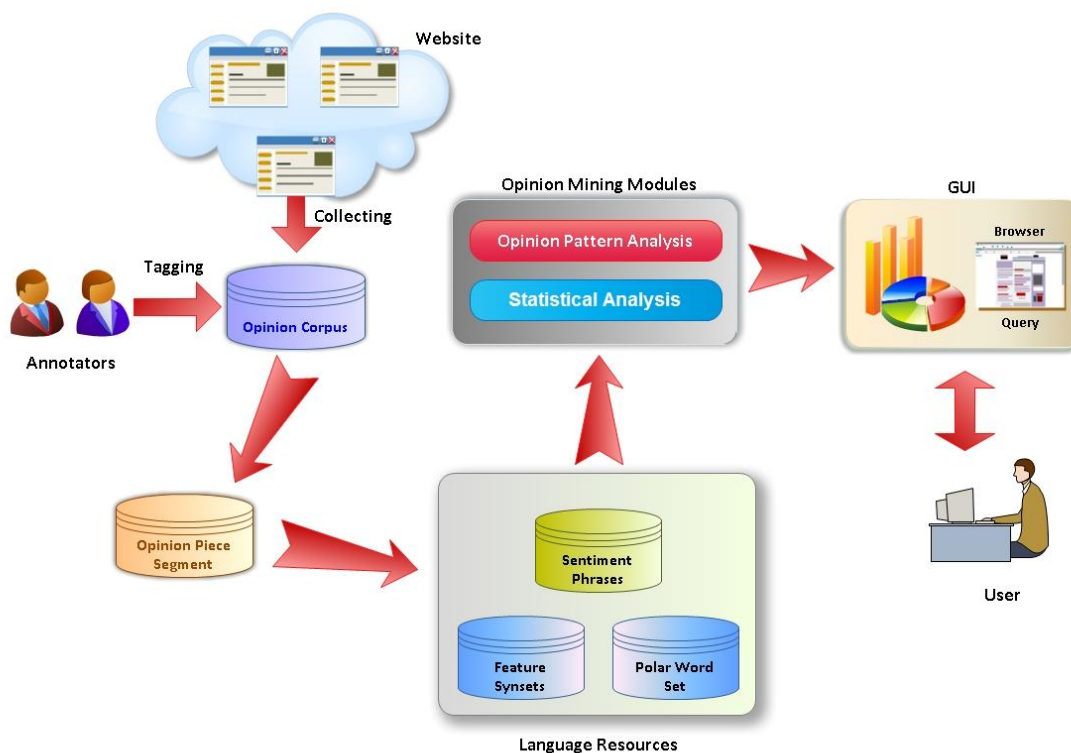
เปรียบเทียบคุณลักษณะโรงแรม

การบริการ	อาหาร	สภาพโรงแรม	สถานที่ตั้ง
ห้องพัก	บริการเพิ่มเติม	ราคา	
การบริการ			
Sentence	Polarity		
ราคาพนักงานบริการดีมาก ดูความคิดเห็นเดิม	+		
การบริการของพนักงานน่ารักมาก ดูความคิดเห็นเดิม	+		
พนักงานเป็นกันเอง ดูความคิดเห็นเดิม	+		
บริการประทับใจ ดูความคิดเห็นเดิม	+		
พนักงานบริการดีมาก ดูความคิดเห็นเดิม	+		
บริการมากกว่าที่นี้ลงตัวและน่าอยู่ ดูความคิดเห็นเดิม	+		
พนักงานสุภาพมากๆ ดูความคิดเห็นเดิม	+		
ที่พักที่ปลอดภัย ดูความคิดเห็นเดิม	+		

รูปที่ 3.14 ข้อความแสดงความคิดเห็นที่นำมาสรุปเป็นข้อมูลทางสถิติ

โครงสร้างของซอฟต์แวร์

ระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรม ประกอบด้วยขั้นตอนการกำกับคลังข้อมูล(Corpus annotation) จากคลังข้อความความคิดเห็น (Opinion Corpus) เพื่อนำชิ้นส่วนข้อความแสดงความคิดเห็น (Opinion Piece Segment) มาแยกเป็นวลี และคำแต่ละประเภท และขั้นตอนการวิเคราะห์เพื่อระบุคุณลักษณะที่ข้อความแสดงความคิดเห็นกล่าวถึง ระบุข้อความความคิดเห็น และแสดงผลสรุปข้อความแสดงความคิดเห็นให้ผู้ใช้ ดังโครงสร้างในรูปที่ 3.15 [4]



รูปที่ 3.15 โครงสร้างระบบวิเคราะห์ข้อความแสดงความคิดเห็น

- ที่มาของไลบรารีและชุดทดสอบ
 - Mysql Connector จากเว็บไซต์ <http://www.mysql.com/products/connector>
 - Java Excel API จากเว็บไซต์ <http://jexcelapi.sourceforge.net>
 - LexToPlus จาก หน่วยปฏิบัติการวิจัยวิทยาการมนุษยภาษา ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC)
 - JavaScript InfoVis Toolkit (thejit) จากเว็บไซต์ <http://thejit.org>
 - ชุดข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรมที่ใช้ในการประเมินผล จากเว็บไซต์ <http://www.agoda.co.th>

3.4 ขอบเขตและข้อจำกัด

1. ระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรมนำข้อความแสดงความคิดเห็นภาษาไทยจากลูกค้าในเว็บไซต์ <http://www.agoda.co.th> ที่แสดงความคิดเห็นเกี่ยวกับโรงแรมในกรุงเทพมหานคร, เชียงใหม่, เชียงราย, ภูเก็ต, ระยอง, หัวหิน, ชะอำ, พัทยา, เกาะสมุย, เกาะช้างและเกาะพีพี เนื่องจากเป็นจังหวัดและสถานที่ท่องเที่ยวที่ได้รับความนิยมจากนักท่องเที่ยวเป็นอันดับต้นๆของประเทศไทย โดยเลือกมาจากจังหวัดและสถานที่ 10 อันดับแรกที่มีข้อความแสดงความคิดเห็นมากที่สุดจากเว็บไซต์ <http://www.agoda.co.th>

2. ระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรมจะวิเคราะห์เฉพาะความคิดเห็นแบบตรงในคุณลักษณะที่กำหนด 7 คุณลักษณะได้แก่ การบริการ, อาหาร, สภาพของโรงแรม, สถานที่ตั้ง, ห้องพัก, สิ่งอำนวยความสะดวก และราคา [5]

3. ระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรมไม่สามารถวิเคราะห์ข้อความแสดงความคิดเห็นเชิงเปรียบเทียบได้ เช่น

- พนักงานบริการระดับเดียวกันกับโรงแรมเอ

จากข้อความแสดงความคิดเห็นนี้ ระบบไม่สามารถตัดสินได้ว่าโรงแรมเอบริการดีหรือไม่ดี จึงไม่สามารถสรุปผลให้ผู้ใช้ได้ว่า ข้อความแสดงความคิดเห็นนี้เป็นขั้วบวกหรือขั้วลบ

- บรรยากาศเหมือนอยู่มัลดีฟส์

จากข้อความแสดงความคิดเห็นดังกล่าว ระบบไม่มีข้อมูลว่าบรรยากาศที่มัลดีฟส์ดีหรือไม่ดี จึงไม่สามารถสรุปผลให้ผู้ใช้ได้ว่า ข้อความแสดงความคิดเห็นนี้เป็นขั้วบวกหรือขั้วลบ

4. ระบบวิเคราะห์ข้อความแสดงความคิดเห็นไม่สามารถวิเคราะห์ข้อความแสดงความคิดเห็นที่ต้องอาศัยการตีความหรือแปลความหมายโดยนัยได้ เช่น

- ถึงโรงแรมจะขนาดเล็ก แต่ราคาไม่เล็กตามขนาด

- บริการเราอย่างกับไปขอพักฟรีอย่างนั้นแหละ

5. โครงการนี้ไม่ได้พัฒนาในส่วนของระบบการตัดคำภาษาไทย ระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรมจึงอ้างอิงผลลัพธ์การตัดคำภาษาไทยจากโปรแกรม LexToPlus ในการวิเคราะห์ข้อความแสดงความคิดเห็น

บทที่ 4 กลุ่มผู้ใช้ระบบ

ผู้ที่ต้องการข้อมูลสรุปจากข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม เพราะระบบจะช่วยวิเคราะห์ข้อความแสดงความคิดเห็น จากข้อความแสดงความคิดเห็นที่เป็นภาษาธรรมชาติออกมาแสดงผลให้ผู้ใช้ในรูปแบบกราฟิกที่มาจากการสรุปข้อมูลในเชิงสถิติ อีกทั้งผู้ใช้สามารถเลือกดูข้อความที่ระบบสกัดออกมาได้อีกด้วย

บทที่ 5 ผลของการทดสอบระบบ

จากการประเมินผลการวิเคราะห์ด้วยชุดข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม 600 ข้อความ แบ่งเป็น 6 คุณลักษณะ (การบริการ, อาหาร, สภาพโรงแรม, สถานที่ตั้ง, ห้องพัก, ราคา) จากเว็บไซต์ <http://www.agoda.co.th> แต่ละคุณลักษณะจะวิเคราะห์จำนวนข้อความแสดงความคิดเห็นทั้งหมด 100 ข้อความ สรุปได้ดังนี้

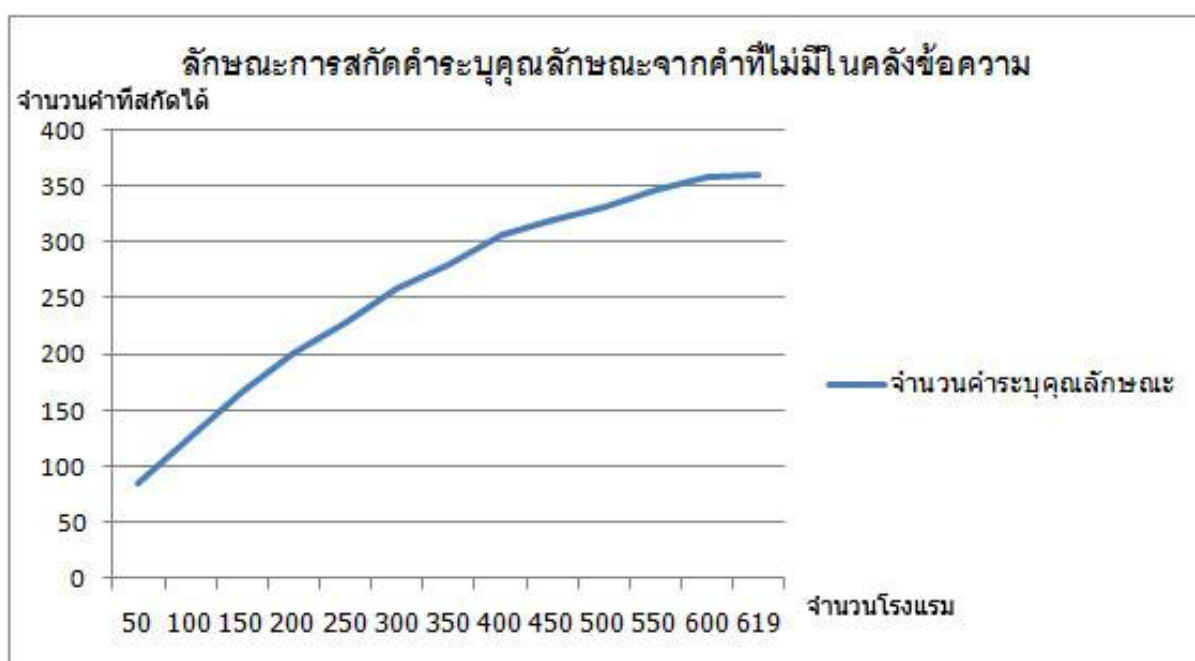
การประเมินผลการสกัดคำระบุคุณลักษณะและคำระบุขั้วจากคำที่ไม่มีในคลังข้อความ

คำระบุคุณลักษณะสามารถสกัดได้ 516 ข้อความจาก 600 ข้อความ

คำระบุขั้วสามารถสกัดได้ 570 ข้อความจาก 600 ข้อความ

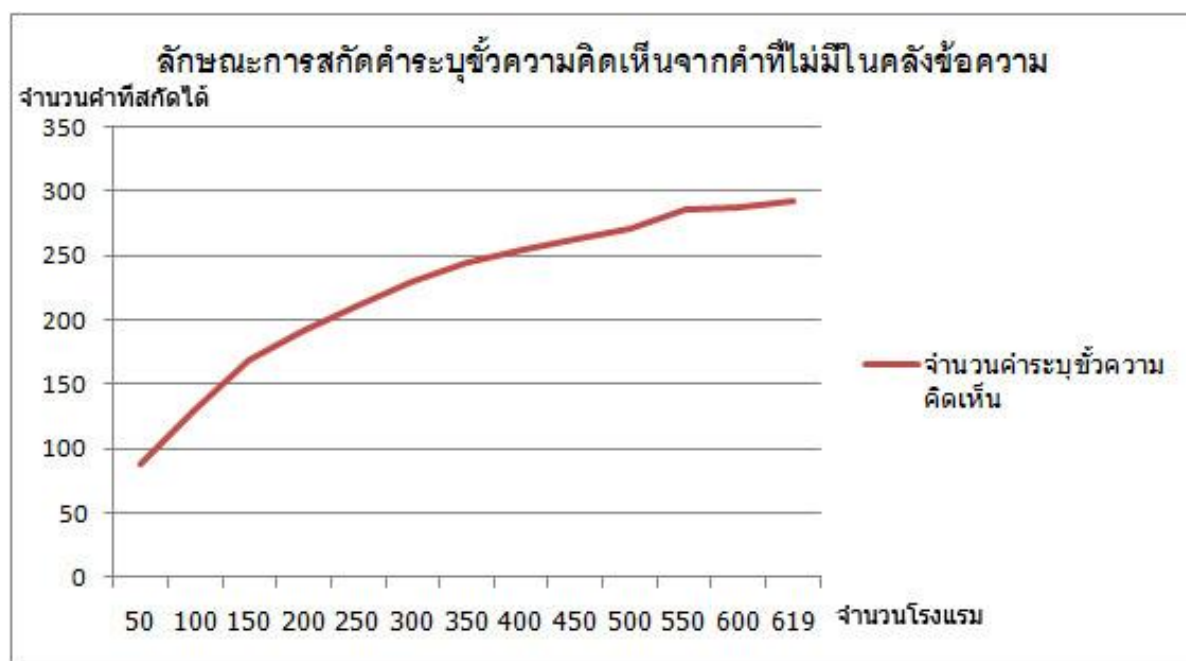
	ความแม่นยำ (Precision)	ความครบถ้วน (Recall)	F-Measure
การสกัดคำระบุคุณลักษณะ (Feature Word Extraction)	ได้รับคะแนน 396.64 คิดเป็นร้อยละ 76.87	ได้รับคะแนน 431.54 คิดเป็นร้อยละ 83.63	80.11
การสกัดคำระบุขั้วความคิดเห็น (Polar Word Extraction)	ได้รับคะแนน 424.72 คิดเป็นร้อยละ 74.51	ได้รับคะแนน 483.03 คิดเป็นร้อยละ 84.74	79.30

การสกัดคำระบุคุณลักษณะจากคำที่ไม่มีในคลังข้อความ จากข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม 619 โรงแรม มีลักษณะการเพิ่มของค่า ดังในภาพที่ 5.1



รูปที่ 5.1 ลักษณะการสกัดคำระบุคุณลักษณะจากคำที่ไม่มีในคลังข้อความ

การสกัดคำระบุข้อความที่เห็นจากคำที่ไม่มีในคลังข้อความ จากข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม 619 โรงแรม มีลักษณะการเพิ่มของคำ ดังในภาพที่ 5.2



รูปที่ 5.2 ลักษณะการสกัดคำระบุข้อความที่เห็นจากคำที่ไม่มีในคลังข้อความ

การประเมินผลการสกัดข้อความ (Information Extraction)

สามารถสกัดได้ 590 ข้อความจาก 600 ข้อความ

	ความแม่นยำ (Precision)	ความครบถ้วน (Recall)	F-Measure
การสกัดคำระบุคุณลักษณะ (Feature Word Extraction)	ได้รับคะแนน 501.82 คิดเป็นร้อยละ 85.05	ได้รับคะแนน 536.85 คิดเป็นร้อยละ 90.99	87.92
การสกัดคำระบุข้อความที่เห็น (Polar Word Extraction)	ได้รับคะแนน 475.74 คิดเป็นร้อยละ 80.63	ได้รับคะแนน 532.96 คิดเป็นร้อยละ 90.33	85.20

จากผลการประเมินข้างต้นสามารถสรุปได้ว่า ระบบวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรมสามารถวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรมจากข้อความแสดงความคิดเห็นด้วยภาษาธรรมชาติได้ จากนั้นสามารถนำข้อความที่วิเคราะห์มาสรุปเป็นข้อมูลเชิงสถิติแล้วแสดงในรูปแบบกราฟิก

บทที่ 6 ปัญหาและอุปสรรค

1. การตัดคำและการสกัดข้อความแสดงความคิดเห็นยังมีข้อผิดพลาด เนื่องจากคำศัพท์บางคำ อาจจะไม่อยู่ในฐานข้อมูลของตัวตัดคำและคลังข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม

แนวทางการแก้ไข สกัดคำระบุคุณลักษณะ และคำระบุข้อเพื่อระบุคำที่ยังไม่มีในคลังข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม เพื่อนำคำไปเพิ่มในคลังข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม

2. ระบบวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรมใช้ไลบรารีและเครื่องมือหลายชนิด ทำให้มีปัญหาในการส่งข้อมูลติดต่อกัน

แนวทางการแก้ไข ใช้ฐานข้อมูลเป็นตัวกลางในการแลกเปลี่ยนข้อมูล

3. ข้อมูลโรงแรมจากเว็บไซต์ <http://www.agoda.co.th> มีเป็นจำนวนมาก ทำให้การดึงข้อมูลใน ส่วนของเทคนิคจินตทัศน์ค่อนข้างช้า และทำให้ต้องใช้พื้นที่หน้าจอในการแสดงผลเกินกว่าที่กำหนด

แนวทางการแก้ไข ลดปริมาณการแสดงผลข้อมูลของเทคนิคจินตทัศน์ลง โดยแบ่งการแสดงผลเป็นส่วนๆของแต่ละโรงแรม

บทที่ 7 แนวทางในการพัฒนาและประยุกต์ใช้ร่วมกับงานอื่นๆ ในขั้นต่อไป

แนวทางในการพัฒนาและประยุกต์ใช้ร่วมกับงานอื่นๆ ได้แก่การนำระบบวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรมนี้ประยุกต์ใช้กับการวิเคราะห์ข้อความแสดงความคิดเห็นในระดับคุณลักษณะสำหรับหัวข้ออื่นๆที่สนใจ ตัวอย่างเช่น การวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับบริการหลังการขายของบริษัท, การวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับคุณภาพสินค้าของบริษัท เป็นต้น ซึ่งในส่วนนี้สามารถที่จะพัฒนาต่อไปได้ในอนาคต เพื่อให้รองรับกับข้อความแสดงความคิดเห็นในหัวข้อที่ผู้ใช้งานต้องการ

บทที่ 8 ข้อสรุปและข้อเสนอแนะ

ระบบวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรมสามารถที่จะวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรมได้ แต่เนื่องจากข้อความแสดงความคิดเห็นภาษาไทย ไม่มีรูปแบบที่แน่นอน จึงทำให้ยังมีข้อผิดพลาดอยู่ หากมีการปรับปรุงหรือแก้ไขเพิ่มเติมก็จะสามารถทำให้ระบบวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรมสามารถวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรมได้อย่างแม่นยำยิ่งขึ้น และสามารถนำไปปรับใช้กับการวิเคราะห์ข้อความแสดงความคิดเห็นในหัวข้ออื่นที่น่าสนใจได้อย่างมีประสิทธิภาพ

บทที่ 9 เอกสารอ้างอิง

- [1] Alisa Kongtho, et al. (2010). **Thailand's Tourism Information Service Based on Semantic Search and Opinion Mining**. Human Language Technology Laboratory (HLT), National Electronics and Computer Technology Center (NECTEC), Thailand
- [2] Bing Liu. **Opinion Mining and Summarization (including review spam detection)**, tutorial given at WWW-2008, April 21, 2008 in Beijing, China
- [3] Choochart Haruechaiyasak, et al. (2010). **Constructing Thai Opinion Mining Resource: A Case Study on Hotel Reviews**. Human Language Technology Laboratory (HLT), National Electronics and Computer Technology Center (NECTEC), Thailand
- [4] D plus Magazine. **ขยับจาก Brand Monitoring สู่ Sentiment Analysis**. [Online]. 2549. ค้นหาคำเมื่อ 10 สิงหาคม 2553. จาก <http://www.dplusmag.com/world-wide-web-social-network/ขยับจาก-brand-monitoring-สู่-sentiment-analysis.html>
- [5] Google. **Visualization: Bar Chart**. [Online]. ค้นหาคำเมื่อ 13 สิงหาคม 2553. จาก <http://code.google.com/intl/th-TH/apis/visualization/documentation/gallery/barchart.html>
- [6] Jennifer Golbeck, Chang Hu. (2009). **Impact of Visualization Methods on Interaction with Search Results**. Human-Computer Interaction Lab University of Maryland, College Park, MD
- [7] Tai Parinya. **Natural Language Processing**. [Online]. ค้นหาคำเมื่อ 30 ธันวาคม 2553 จาก <http://www.tai-parinya.com/node/269>
- [8] veer. **เปิดโลก Ajax ตอน JSON**. [Online]. 2549. ค้นหาคำเมื่อ 9 พฤษภาคม 2553. จาก <http://veer.exteen.com/20060110/ajax-json>
- [9] ทองจุล ชันขาว. **การใช้ Regular Expression ใน PHP**. [Online]. ค้นหาคำเมื่อ 2 พฤศจิกายน 2553 จาก <http://www.mindphp.com/modules.php?name=Content&pa=showpage&pid=78>
- [10] ปัทมา ว่าพัฒน์วงศ์. **จำนวนโรงแรมและห้องพักที่สามารถรองรับนักท่องเที่ยว**. [Online]. ค้นหาคำเมื่อ 10 สิงหาคม 2553. จาก <http://www.popterms.mahidol.ac.th/newsletter/showarticle.php?articleid=102>

บทที่ 10 ประวัติส่วนตัวและผลงานดีเด่นของผู้พัฒนา



นางสาวรัญญา วรรณศรี

สถานที่ติดต่อ

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

โทรศัพท์

08-5002-2231

E-mail

vanness_a@hotmail.com

ประวัติดีเด่น

- ผู้ช่วยวิทยากรค่ายยุวคอมพิวเตอร์ ประจำปี2550 โดยสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.)
- ผู้ช่วยวิทยากรค่ายยุวคอมพิวเตอร์ ประจำปี2551 โดยสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.)
- เข้าร่วมการแข่งขัน ACM ICPC THAILAND 2009 โดยภาควิชาวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตภูเก็ต
- วิทยากรค่ายยุวคอมพิวเตอร์ ประจำปี2552 โดยสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.)
- เป็นนักศึกษาผู้ช่วยสอนโครงการอบรมคอมพิวเตอร์โอลิมปิก ปี 2553 (ค่าย2) ณ มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต
- เข้าร่วมการแข่งขัน ACM-ICPC Thailand Central Region Programming Contest 2010 โดยภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ร่วมกับสำนักงานส่งเสริมอุตสาหกรรมซอฟต์แวร์แห่งชาติ (องค์การมหาชน)

บทที่ 11 ภาคผนวก

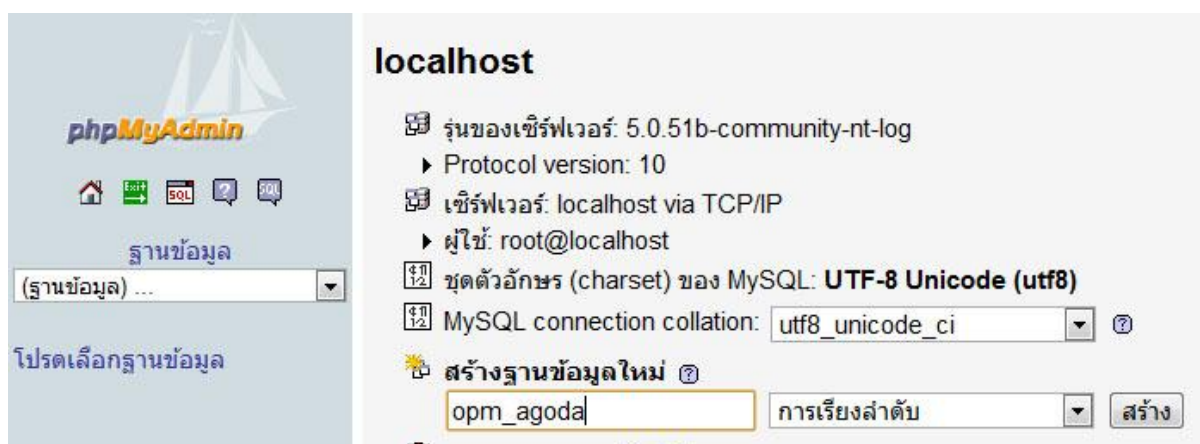
11.1 คู่มือการติดตั้ง

1) ติดตั้งโปรแกรม AppServ ในรูปที่ 11.1 เพื่อใช้ฐานข้อมูล MySQL กำหนด user เป็น root password เป็น 3101



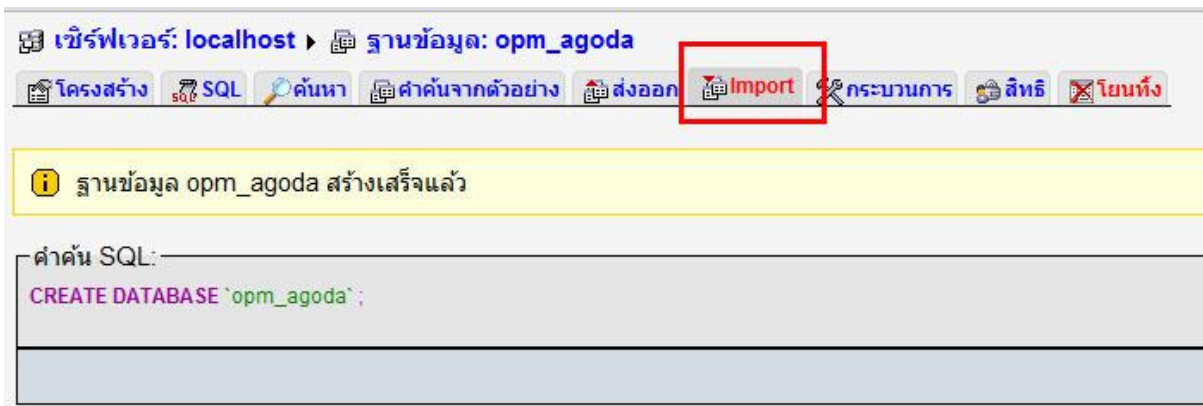
รูปที่ 11.1 โปรแกรม AppServe

จากนั้นเปิด Web Browser ขึ้นมา ใส่ address เป็น localhost เพื่อเปิด phpMyAdmin สร้างฐานข้อมูลชื่อ “opm_agoda” ดังปรากฏในรูปที่ 11.2

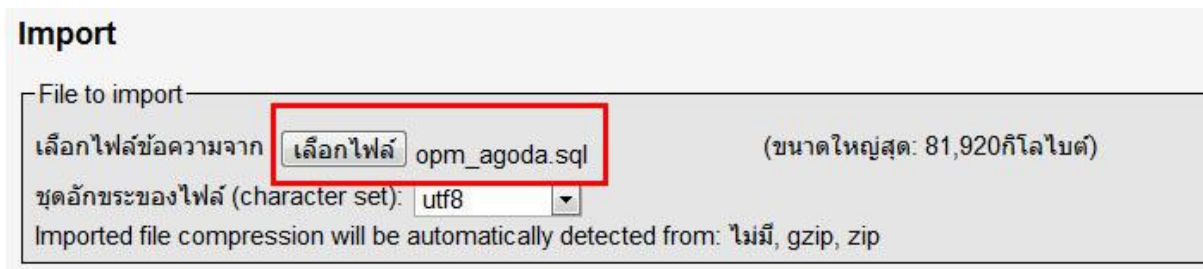


รูปที่ 11.2 การสร้างฐานข้อมูล opm_agoda

เมื่อสร้างฐานข้อมูลเสร็จแล้ว คลิกเลือก Import ตามรูปที่ 11.3 จากนั้นเลือกไฟล์ opm_agoda.sql เพื่อนำตารางเข้าสู่ฐานข้อมูลที่สร้าง ดังในรูปที่ 11.4

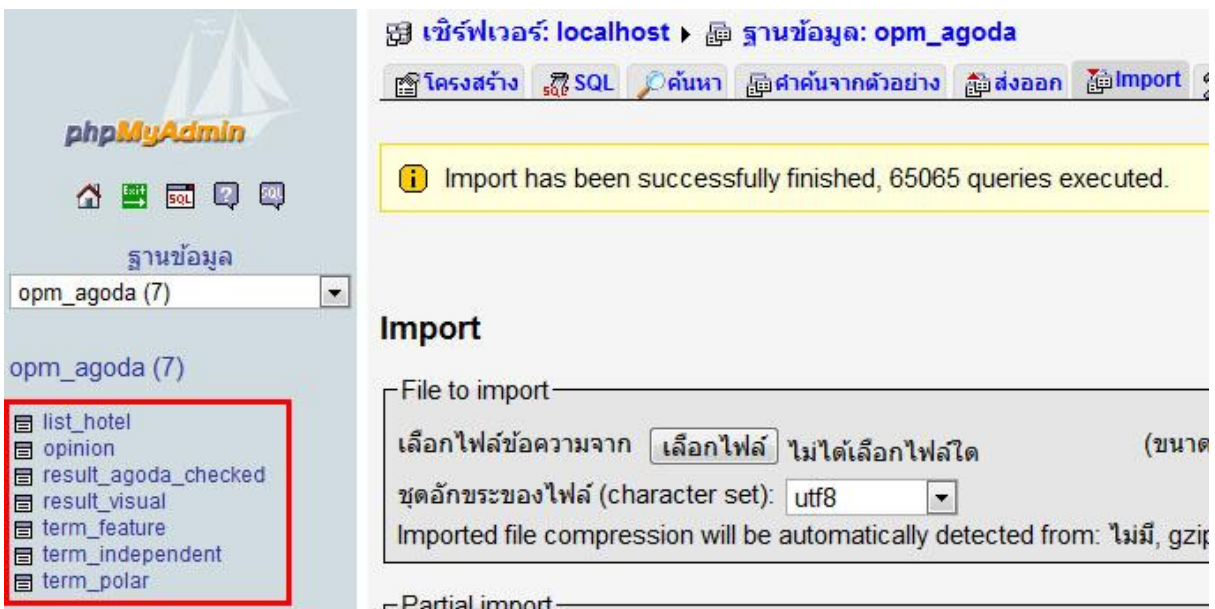


รูปที่ 11.3 การเลือกการนำเข้าข้อมูลสู่ฐานข้อมูล



รูปที่ 11.4 การนำตารางเข้าสู่ฐานข้อมูล opm_agoda

เมื่อการนำตารางเข้าสู่ฐานข้อมูลเสร็จเรียบร้อยแล้ว จะปรากฏรายชื่อตารางที่คอลัมน์ซ้ายมือ และข้อความแสดงสถานะการทำงานทางคอลัมน์ขวามือ ดังในรูปที่ 11.5



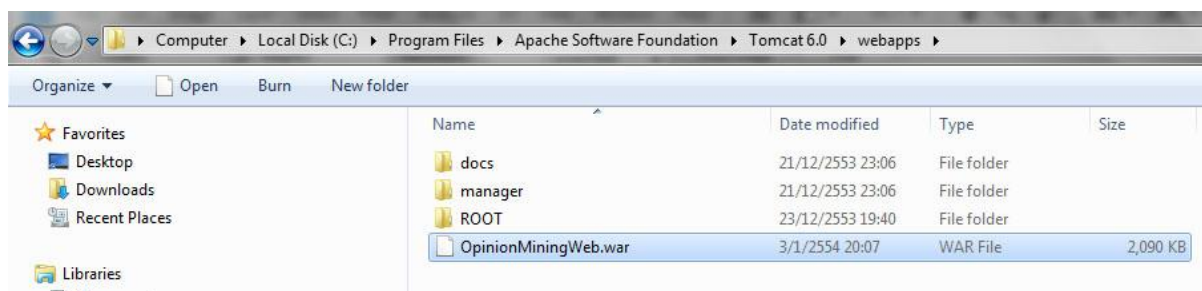
รูปที่ 11.5 หน้าจอแสดงสถานะการนำเข้าตารางสู่ฐานข้อมูล

2) ติดตั้งโปรแกรม Apache Tomcat 6.0 ในรูปที่ 11.6 เพื่อใช้เป็นเซิร์ฟเวอร์จำลองกายเครื่อง



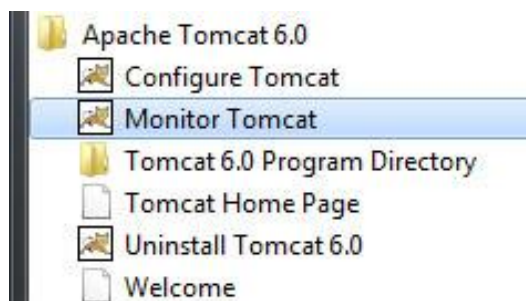
รูปที่ 11.6 โปรแกรม Apache Tomcat 6.0

3) คัดลอกไฟล์ OpinionMiningWeb.war ไปไว้ที่ Folder webapps ดังในรูปที่ 11.7 โดยมี Directories Path เป็น Folder ที่ติดตั้ง Apache Tomcat 6.0 ซึ่งก็คือ Folder Apache Software Foundation > Tomcat6.0 > webapps



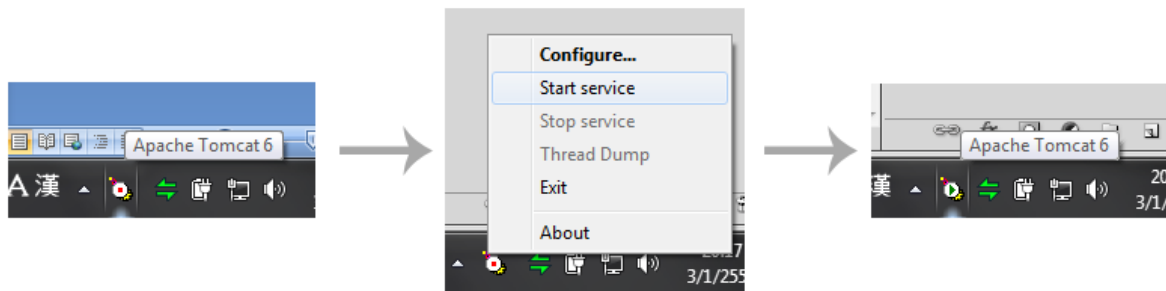
รูปที่ 11.7 คัดลอกไฟล์ OpinionMiningWeb.war วางไว้ที่ Folder webapps

4) คลิกเปิดโปรแกรม Monitor Tomcat ตัวอย่างในรูปที่ 11.8



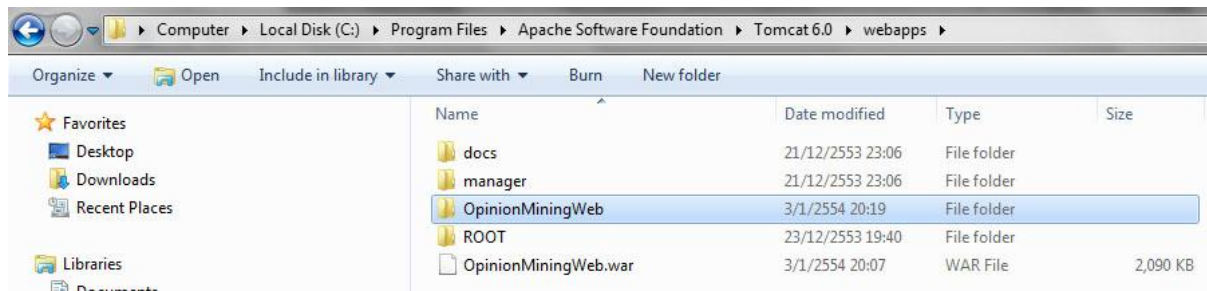
รูปที่ 11.8 คลิกเปิดโปรแกรม Monitor Tomcat

หลังจากนั้นที่ไอคอนมุมขวาล่าง เลือก Start Service ไอคอนจะเปลี่ยนจากสีเหลี่ยมสีแดงเป็น สีส้มเหลี่ยมสี่เหลี่ยม ดังในรูปที่ 11.9



รูปที่ 11.9 การ Start Service

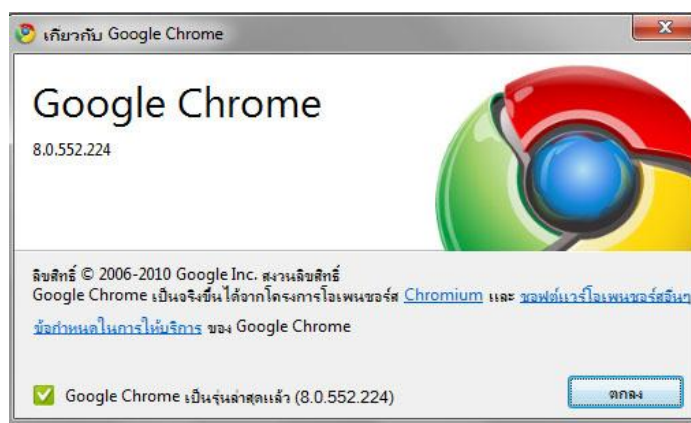
5) เมื่อ Start Service แล้วจะเห็นว่าใน Folder webapps มี Folder ชื่อ OpinionMiningWeb เพิ่มขึ้น ดังในรูปที่ 11.10



รูปที่ 11.10 Folder OpinionMiningWeb ปรากฏขึ้นใน Folder webapps

11.2 คู่มือการใช้งาน

- 1) เปิด Web Browser ขึ้นมา ในที่นี้ใช้ Google Chrome รุ่น 8.0.552.224 ดังในรูปที่ 11.11



รูปที่ 11.11 Web Browser ที่ใช้ Google Chrome รุ่น 8.0.552.224

- 2) พิมพ์ที่อยู่ localhost:8080/OpinionMiningWeb ที่ address bar ดังในรูปที่ 11.12 กดยืนยัน เพื่อแสดงหน้าเว็บไซต์



รูปที่ 11.12 การเข้าถึงหน้าเว็บไซต์

3) เลือกสถานที่ตั้งของโรงแรมเพื่อดูข้อมูลการวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม เหมือนในตัวอย่างรูปที่ 11.13

รูปที่ 11. 13 การเลือกสถานที่ตั้งของโรงแรม

4) เลือกโรงแรมที่ต้องการดูข้อมูลการวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม เหมือนในตัวอย่างรูปที่ 11.14 ระบบจะแสดงข้อมูลที่เป็นกราฟิกเป็นค่าเริ่มต้น



รูปที่ 11. 14 การเลือกโรงแรมที่ต้องการดูข้อมูลการวิเคราะห์ข้อความแสดงความคิดเห็น

5) คลิกที่ “แสดงผลแบบข้อความ” เพื่อดูข้อมูลการวิเคราะห์ข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรมที่ถูกประมวลผลเพื่อแสดงเป็นแบบกราฟิก ดังในรูปที่ 11.15

Sentence	Polarity
ราคาพนักงานบริการดีมาก <u>ดูความคิดเห็นเดิม</u>	+
การบริการของพนักงานน่ารักมาก <u>ดูความคิดเห็นเดิม</u>	+
พนักงานเป็นกันเอง <u>ดูความคิดเห็นเดิม</u>	+
บริการประทับใจ <u>ดูความคิดเห็นเดิม</u>	+
พนักงานบริการดีมาก <u>ดูความคิดเห็นเดิม</u>	+
บริการมากกว่าที่นี้ลงตัวและน่าอยู่ <u>ดูความคิดเห็นเดิม</u>	+
พนักงานสุภาพมาก <u>ดูความคิดเห็นเดิม</u>	+
ที่พักรูปอลอดภัย <u>ดูความคิดเห็นเดิม</u>	-

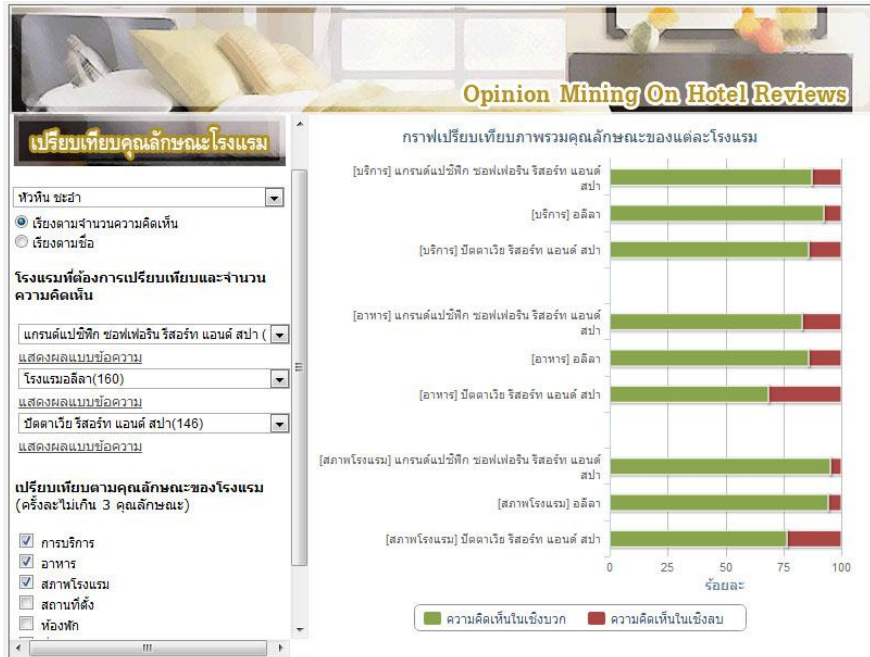
รูปที่ 11. 15 การเลือกโรงแรมที่ต้องการเพื่อดูข้อความที่นำมาประมวลผลเพื่อแสดงเป็นแบบกราฟิก

6) เลือกรการเน้นสีคำที่ระบุคุณลักษณะ หรือคำระบุข้อความความคิดเห็นได้ด้วยการเลือกที่ Check Box “เลือกแสดงคำระบุคุณลักษณะ” และ “เลือกแสดงคำระบุข้อความความคิดเห็น” ดังในรูปที่ 11.16

Sentence	Polarity
ราคาพนักงานบริการดีมาก <u>ดูความคิดเห็นเดิม</u>	+
การบริการของพนักงานน่ารักมาก <u>ดูความคิดเห็นเดิม</u>	+
พนักงานเป็นกันเอง <u>ดูความคิดเห็นเดิม</u>	+
บริการประทับใจ <u>ดูความคิดเห็นเดิม</u>	+
พนักงานบริการดีมาก <u>ดูความคิดเห็นเดิม</u>	+
บริการมากกว่าที่นี้ลงตัวและน่าอยู่ <u>ดูความคิดเห็นเดิม</u>	+
พนักงานสุภาพมาก <u>ดูความคิดเห็นเดิม</u>	+
ที่พักรูปอลอดภัย <u>ดูความคิดเห็นเดิม</u>	-

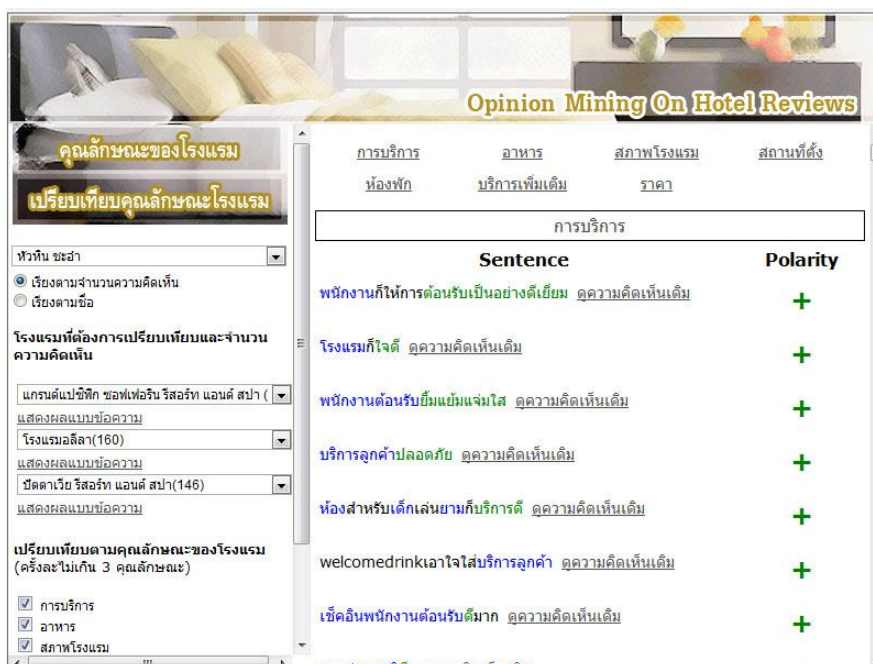
รูปที่ 11. 16 การเลือกรการเน้นสีคำที่ระบุคุณลักษณะ หรือคำระบุข้อความความคิดเห็น

6) เลือกโรงแรม 3 โรงแรมในสถานที่ที่ต้องการมาเปรียบเทียบตามแต่ละคุณลักษณะได้ ไม่เกิน 3 คุณลักษณะ ดังตัวอย่างในรูปที่ 11.17 โดยแสดงในรูปแบบกราฟิกเป็นค่าเริ่มต้น



รูปที่ 11. 17 การเลือกโรงแรมเพื่อเปรียบเทียบในแต่ละคุณลักษณะ

7) เลือกโรงแรม 3 โรงแรมในสถานที่ที่ต้องการมาเปรียบเทียบตามแต่ละคุณลักษณะได้ ไม่เกิน 3 คุณลักษณะ และสามารถเลือกดูข้อความแสดงความคิดเห็นของแต่ละโรงแรมได้ ด้วยการคลิกที่ “แสดงผลแบบข้อความ” ที่อยู่ใต้ Drop down list ของแต่ละโรงแรมที่เลือก ดังแสดงในรูปที่ 11.18



รูปที่ 11. 18 การแสดงข้อความแสดงความคิดเห็นของโรงแรมที่ถูกเปรียบเทียบในแต่ละคุณลักษณะ