

การวิเคราะห์ความคิดเห็นเกี่ยวกับร้านอาหารบนเว็บไซต์รีวิว

Sentiment Analysis of Restaurant Reviews on Review Web Sites

ปราชญ์ภาคย์ เหล่าสังข์สุข¹ อนัส จินดา² และสิริยา สิทธิสาร^{3*}
Prachayaphak Laowsungsk¹, Anus Jinda² and Siraya Sitthisarn^{3*}

บทคัดย่อ

ในปัจจุบันจำนวนร้านอาหารมีมากมายและหลากหลายทั้งประเภทอาหารและการบริการทำให้การตัดสินใจของผู้บริโภคในการเลือกใช้บริการร้านอาหารยากขึ้น ผู้บริโภคจึงต้องการข้อมูลเพื่อช่วยในการตัดสินใจ และช่องทางหนึ่งที่ผู้คนในโลกออนไลน์นิยมใช้คือการเข้าไปหาข้อมูลในเว็บไซต์รีวิวออนไลน์ อย่างไรก็ตามเนื่องจากความคิดเห็นส่วนใหญ่เป็นข้อความทำให้ผู้ใช้เว็บไซต์ดังกล่าวเสียเวลาในการอ่านและหาข้อสรุปเพื่อประกอบการตัดสินใจ ในส่วนของผู้ประกอบการเองก็ยากในการหาข้อดีข้อเสียสำหรับปรับปรุงร้านให้ดีขึ้น ดังนั้นงานวิจัยนี้จึงเสนอระบบวิเคราะห์และสรุปความคิดเห็นเกี่ยวกับร้านอาหารจากเว็บไซต์รีวิวโดยอัตโนมัติ การวิเคราะห์จะใช้เทคนิคต่างๆ ในการประมวลผลภาษาธรรมชาติ เช่น เทคนิคการตัดคำ การวิเคราะห์ประเภทคำและรูปประโยค เพื่อหาความหมายเชิงบวกหรือเชิงลบของประโยค แล้วจึงคำนวณและแสดงผลออกมาเป็นค่าระดับความพึงพอใจสำหรับคุณลักษณะต่างๆในรูปแบบกราฟิกทั้งนี้เพื่อให้ผู้ใช้เข้าใจได้ง่าย ซึ่งจากผลการประเมินความพึงพอใจของผู้ใช้ในด้านความสอดคล้องของผลสรุปจากระบบและผลสรุปจากผู้ใช้พบว่าการสรุปความคิดเห็นด้านบริการและอาหารอยู่ในระดับดี ขณะที่ด้านอื่นๆ อยู่ในระดับพอใช้

คำสำคัญ: การวิเคราะห์ความคิดเห็น, ความคิดเห็นเกี่ยวกับร้านอาหาร, การสรุปความคิดเห็น

Abstract

Review websites have widely used for searching restaurant information that enables user decision supporting. Unfortunately, the review comments are text messages. Thus, users have to read all comments of the interested restaurants and then summarize some issues. This process is time consuming. Restaurant managers also found that it is difficult to extract the customers' critics from these comments. This paper proposes a feature-based sentiment analysis system for automatic summarizing customers' comments. Natural Language Processing (NLP) techniques, including tokenization, analyzing part of speech and sentence pattern analysis, were used for determining the polarity of sentences. Finally a polarity score of each feature was computed and displayed in graphic visualization. The user evaluation was conducted to find user satisfaction on correctness of the sentiment analysis. The result shows that users provide high satisfaction on the analysis for the food and service features, while others reach the fair level of satisfaction.

Keywords: Sentiment Analysis, Restaurant Reviews, Opinion Summarization

^{1,2} นิสิต สาขาวิชาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยทักษิณ พัทลุง 93210

³ ผศ., สาขาวิชาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยทักษิณ พัทลุง 93210

^{1,2} Undergraduate students, Department of Computer and Information Technology, Faculty of Science, Thaksin University, Phatthalung, 93210

³ Assistant Professor, Department of Computer and Information Technology, Faculty of Science, Thaksin University, Phatthalung, 93210

* Corresponding author: E-mail address: ssitthisarn@gmail.com

บทนำ

ในปัจจุบันการรับประทานอาหารนอกบ้านโดยเฉพาะในภัตตาคารมีราคาค่อนข้างแพง ดังนั้นผู้บริโภคส่วนใหญ่ย่อมมีความต้องการบริการที่ดีและคุณภาพอาหารที่คุ้มค่ากับเงินที่จ่ายไป ช่องทางหนึ่งที่ผู้บริโภคจะทราบถึงข้อมูลร้านอาหารเหล่านั้นก็โดยการหาข้อมูลในเว็บไซต์ทางสังคมที่ให้บริการรีวิวร้านอาหาร เช่น เว็บไซต์ดั่งใน (<http://www.wongnai.com>) และ TripAdvisor (<http://tripadvisor.com>) เป็นต้น เว็บไซต์วีวส่วนใหญ่นำเสนอความคิดเห็นจากผู้บริโภคที่เคยใช้บริการร้านอาหารในรูปแบบข้อความซึ่งถือเป็นข้อมูลที่ไม่มีโครงสร้าง หากร้านอาหารใดมีผู้แสดงความคิดเห็นมากก็จะปรากฏข้อความหลายร้อยข้อความ ทำให้ผู้ใช้ต้องเสียเวลาอ่านข้อคิดเห็นทั้งหมดแล้วทำการสรุปข้อดีข้อเสียในด้านต่างๆ สำหรับร้านที่ตนสนใจ แม้ว่าบางเว็บไซต์จะทราบถึงข้อจำกัดในเรื่องนี้และได้สร้างฟังก์ชัน rating เพื่อให้ผู้ใช้ให้คะแนนร้านอาหารในคุณลักษณะด้านต่างๆ แต่ความยุ่งยากนั้นก็กลับทำให้ผู้ใช้เข้ามาแสดงความคิดเห็นมีจำนวนลดลง สุดท้ายหลายๆ เว็บจึงมีเพียงกล่องข้อความให้ผู้รีวิวเขียนข้อความ อัปโหลดรูปภาพ และลดการ rating เหลือเฉพาะในภาพรวมของร้านเท่านั้น

จากปัญหาและที่มาข้างต้นผู้วิจัยได้นำเสนอระบบการวิเคราะห์ความคิดเห็นของผู้บริโภคบนเว็บไซต์วีวร้านอาหาร เพื่ออำนวยความสะดวกสำหรับผู้บริโภคในการรับรู้คุณภาพของร้านอาหารในด้านต่างๆ โดยจะแบ่งรายละเอียดการสรุปออกเป็น 5 ข้อคือ (1) รสชาติอาหาร (2) บริการของร้าน (3) บรรยากาศของร้าน (4) ความสะอาด และ (5) ราคาอาหาร ระบบจะวิเคราะห์ข้อความด้วยเทคนิคการตัดคำ การวิเคราะห์ประเภทคำ และรูปประโยค เพื่อหาความหมายเชิงบวกหรือเชิงลบของประโยค แล้วจึงคำนวณและแสดงผลสรุปออกมาเป็นค่าระดับความพึงพอใจสำหรับแต่ละด้านในรูปแบบกราฟิกเพื่อให้ผู้บริโภคใช้ประกอบการตัดสินใจได้ง่ายยิ่งขึ้น นอกจากนี้ผู้พัฒนาคาดว่าผลสรุปจากความคิดเห็นจะเป็นประโยชน์สำหรับผู้ประกอบการร้านอาหารที่สามารถรับรู้ข้อบกพร่องของร้านตนเองเพื่อใช้ในการปรับปรุงร้านให้ดีขึ้น

งานวิจัยที่เกี่ยวข้อง

การวิเคราะห์ความคิดเห็นส่วนใหญ่จะจำแนกคิดเห็นเป็น 3 ข้อคือ ข้อความความคิดเห็นเชิงบวก เชิงลบ และเป็นกลาง ในการวิเคราะห์จะแบ่งได้เป็น 3 ระดับ [1-3] คือ (1) ระดับเอกสาร (Document Level) [4] เป็นการวิเคราะห์ความคิดเห็นทั้งเอกสารเพื่อแยกข้อความความคิดเห็น ยกตัวอย่างเช่น Turney [5] ได้นำเสนอวิธีการนับความถี่ของคำคุณศัพท์ที่แสดงข้อความความคิดเห็นแล้วสรุปว่าความคิดเห็นในเอกสารนั้นเป็นบวกหรือลบ ซึ่งจะเห็นว่าเป็นรูปแบบการวิเคราะห์ที่ไม่ละเอียดพอ (2) ระดับประโยค (Sentence Level) คือ การนำความคิดเห็นมาวิเคราะห์ในระดับประโยคแล้วทำการแยกข้อ เช่น FANALYTICS [6] ได้เสนอการวิเคราะห์ความคิดเห็นของเพื่อน ในเฟสบุ๊คที่มีต่อข้อความที่ผู้ใช้โพสต์แล้วทำการจำแนกความคิดเห็นแบ่งตามจำนวนประโยคข้อบวกหรือลบ (3) ระดับคุณลักษณะ (Feature Level) คือการวิเคราะห์ความคิดเห็นโดยแบ่งแยกตามคุณลักษณะ จากนั้นนำมาวิเคราะห์ว่าผู้แสดงความคิดเห็นมีทัศนคติกับคุณลักษณะนั้นอย่างไร ส่วนมากการวิเคราะห์ความคิดเห็นในระดับคุณลักษณะจะใช้เพื่อวิเคราะห์ความคิดเห็นของผู้ใช้ต่อผลิตภัณฑ์และบริการ เช่น โปรแกรม S-Sense [7] ที่ใช้วิเคราะห์ความพึงพอใจของผู้ใช้ต่อคุณลักษณะด้านต่างๆ ของผลิตภัณฑ์จากความคิดเห็นของผู้คนในสังคมออนไลน์ เป็นต้น

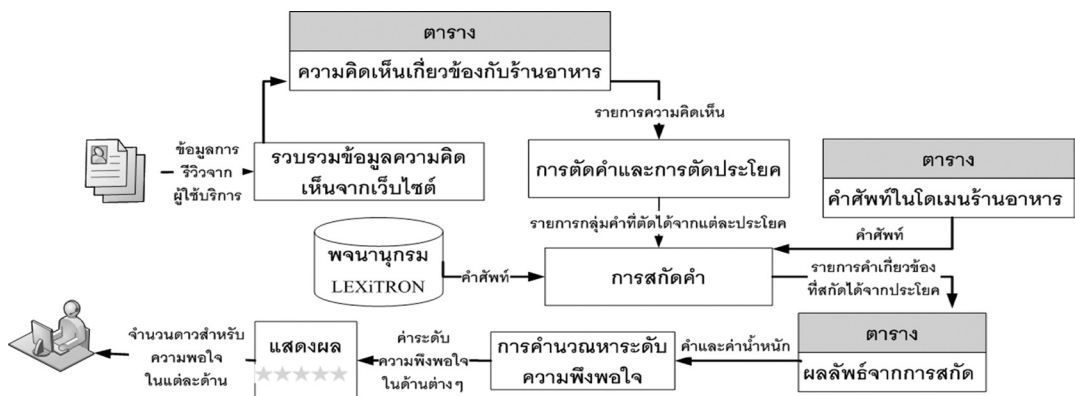
สำหรับเทคนิคการวิเคราะห์ความคิดเห็นแบ่งได้เป็น 2 แนวทางหลักคือ (1) เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) [8] ที่อาศัยเทคนิคการประมวลผลภาษาธรรมชาติและพจนานุกรม แนวทางนี้จะอาศัยคลังคำ (Corpus) ที่รวบรวมจากพจนานุกรมและแหล่งข้อมูลต่างๆ ในโดเมนที่สนใจ คลังคำจะใช้ในการส่วนการตัดคำและการวิเคราะห์ความหมายของคำที่ปรากฏในประโยค เพื่อทำการหาว่าประโยคนั้นเกี่ยวข้องกับคุณลักษณะใด และมีความหมายไปทางขั้วใด งานที่เกี่ยวข้องตามแนวทางนี้ได้แก่ Prombut, N., Temtanapat, Y [9] ได้เสนอ

การวิเคราะห์ความคิดเห็นในสินค้าโทรศัพท์มือถือโดยใช้คลังคำที่รวบรวมจากพจนานุกรมและทำการสร้างฐานข้อมูล คำศัพท์ที่ใช้ใน โดเมน โทรศัพท์มือถือ กฎในการหารูปแบบประโยคตามหลัก ไวยากรณ์จะถูกใช้เพื่อระบุประธาน คำคุณศัพท์ คำบ่งบอกคุณลักษณะและปริมาณ เพื่อคำนวณหาว่าผู้ใช้มีทัศนคติอยู่ในช่วงบวกหรือช่วงลบสำหรับ แต่ละคุณลักษณะ เช่นเดียวกับการวิเคราะห์ความคิดเห็นเกี่ยวกับโรงแรมใน Wonsri, W. [10] ที่ใช้นิพจน์ปรกติ (Regular Expression) ในการจำแนกประเภทของคำในข้อความแสดงความคิดเห็นแล้วนำไปเปรียบเทียบกับรูปแบบ ไวยากรณ์โครงสร้างทางภาษา ก่อนที่จะคำนวณความพึงพอใจจากกลุ่มคำคุณศัพท์และคำบ่งบอกปริมาณในแต่ละคุณลักษณะ ซึ่งความถูกต้องในการวิเคราะห์ความคิดเห็นตามแนวทางนี้จะขึ้นอยู่กับความครอบคลุมของคำศัพท์ ใน โดเมนและเทคนิคในการวิเคราะห์รูปประโยค (2) เทคนิคการเรียนรู้แบบมีผู้สอน (Supervised Learning) [11] ซึ่งอาศัยอัลกอริทึมต่างๆในการทำเหมืองข้อความ เช่น ต้นไม้การตัดสินใจ นาอ็ฟเบย์ เป็นต้น สำหรับเทคนิคใน แนวทางนี้จะทำการแทนคำในข้อความคิดเห็นให้อยู่ในรูปแบบโครงสร้าง เช่น การแทนด้วยถุงคำ (bag-of-word) และ คำ TF-IDF เป็นต้น ทั้งนี้เพื่อใช้ในการสร้างแบบจำลองการวิเคราะห์ข้อมูล โดยจะมีชุดข้อมูลการเรียนรู้ (Training Set) ที่นำไปใช้ในการทำนายรูปแบบของข้อมูลทดสอบ (Test Set) งานวิจัยตามแนวทางนี้ได้แก่ งานของเจตรินทร์ วงศ์ศิลป์ และคณะ [12] ที่เสนอกระบวนการวิเคราะห์ความรู้สึกของนักท่องเที่ยวต่อสถานที่ท่องเที่ยวในภาคอีสานที่ใช้การแทนคำ ด้วยคำ TF-IDF จากนั้นทำการสร้าง โมเดลเพื่อวิเคราะห์ด้วยนาอ็ฟเบย์ โดยคุณภาพความถูกต้องของการวิเคราะห์ ตามแนวทางการเรียนรู้แบบมีผู้สอนนี้จะขึ้นอยู่กับจำนวนและคุณภาพข้อมูลที่ใช้เป็นชุดข้อมูลการเรียนรู้

สำหรับงานวิจัยในบทความนี้จะเป็นการวิเคราะห์ในระดับคุณลักษณะ โดยอาศัยเทคนิคการประมวลผล ภาษาธรรมชาติและคลังศัพท์เช่นเดียวกับ Prombut, N., Temtanapat, Y. and Wonsri, W. [9-10] เนื่องจากผู้วิจัยพบว่า ข้อมูลในเว็บไซต์มีความหลากหลายและการเปลี่ยนแปลงสูงมากถ้าใช้เทคนิคการเรียนรู้แบบมีผู้สอนจะมี ค่าใช้จ่ายในการปรับปรุงโมเดลอยู่เสมอ อย่างไรก็ตามผู้วิจัยได้เพิ่มส่วนการกำกับค่าน้ำหนักของคำคุณศัพท์ใน คลังคำตามกลุ่มคุณลักษณะเพื่อมาเสริม ในการคำนวณระดับค่าความพึงพอใจเชิงบวกและเชิงลบในแต่ละประโยคว่า มีระดับบวกหรือลบที่แตกต่างกันเพียงใด ก่อนที่จะแสดงผลข้อสรุปในรูปแบบกราฟิกให้สอดคล้องกับความคิดเห็นจาก ผู้วิจัยทั้งหมดในภาพรวม

การดำเนินการวิจัย

ระบบวิเคราะห์ความคิดเห็นของผู้บริโภคจากเว็บไซต์วีวร้านอาหารนี้จะสอดคล้องกับสถาปัตยกรรม ในภาพที่ 1 ซึ่งประกอบไปด้วย 5 องค์ประกอบหลัก ดังนี้



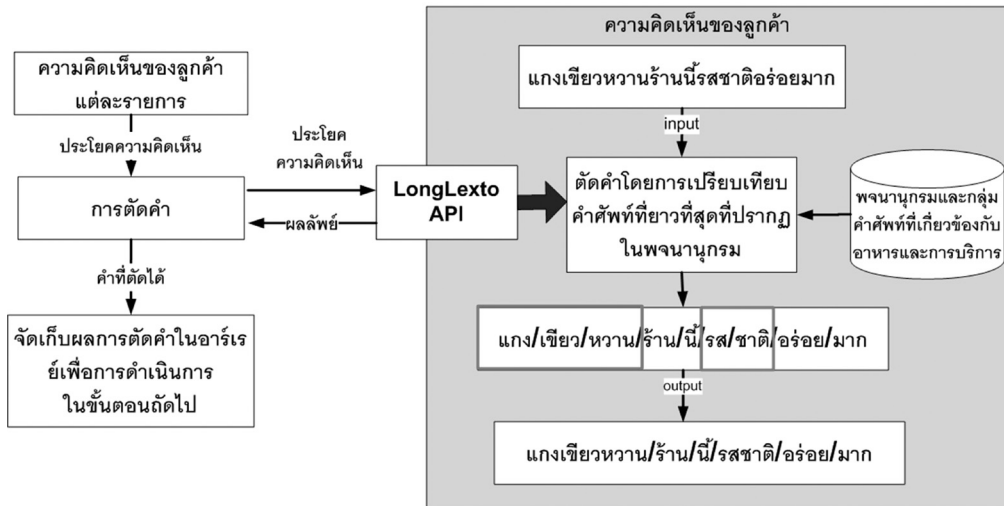
ภาพที่ 1 สถาปัตยกรรมระบบวิเคราะห์ความคิดเห็นของผู้บริโภคจากเว็บไซต์วีวร้านอาหาร

1. การรวบรวมข้อมูลความคิดเห็น

ความคิดเห็นเกี่ยวกับร้านอาหารจากเว็บไซต์วีวจะถูกรวบรวมและจัดเก็บลงในตารางความคิดเห็นเกี่ยวกับร้านอาหาร ที่จะประกอบด้วยแอตทริบิวต์ : รหัสความคิดเห็น ชื่อร้าน และรายละเอียดของความคิดเห็น ตัวอย่างความคิดเห็นได้แก่ “ร้านโกข้าวอาหารกินๆ ไม่ค่อยประทับใจเท่าไร เด็กลีฟไม่คอยสนใจบริการ” เป็นต้น

2. ตัดคำ (Tokenization)

ระบบจะนำประโยคความคิดเห็นแต่ละรายการมาทำการตัดคำโดยใช้เครื่องมือ LongLexto API [13] การตัดคำจะใช้วิธีการตัดคำแบบยาวที่สุดโดยเทียบกับพจนานุกรม LEXiTRON [14] โดยการตัดคำจะเริ่มพิจารณาจากซ้ายไปขวาตามหลักการเขียนภาษาไทย จากนั้นทำการนำ “คำหยุด” (Stop Word) ซึ่งเป็นคำที่ไม่จำเป็นสำหรับการวิเคราะห์ที่ออกเพื่อลดขนาดของข้อมูลที่ใช้ในการประมวลผล จากภาพที่ 2 เมื่อนำประโยคความคิดเห็น “แกงเขียวหวานร้านนี้รสชาติอร่อยมาก” มาตัดคำ ก็จะพบว่าสำหรับ “แกงเขียวหวาน” สามารถตัดคำตามพจนานุกรมได้ 2 แบบ คือคำว่า “แกง” และ คำว่า “แกงเขียวหวาน” แต่เมื่อใช้หลักการตัดคำแบบยาวที่สุด จึงเลือกตัดคำเป็น “แกงเขียวหวาน” เป็นต้น ซึ่งผลลัพธ์ของคำที่ตัดได้เหล่านี้มาเก็บไว้ในตัวแปรอาร์เรย์ชื่อ ArrList เพื่อให้ง่ายต่อขั้นตอนการสกัดคำต่อไป



ภาพที่ 2 ขั้นตอนการตัดคำจากประโยคความคิดเห็น

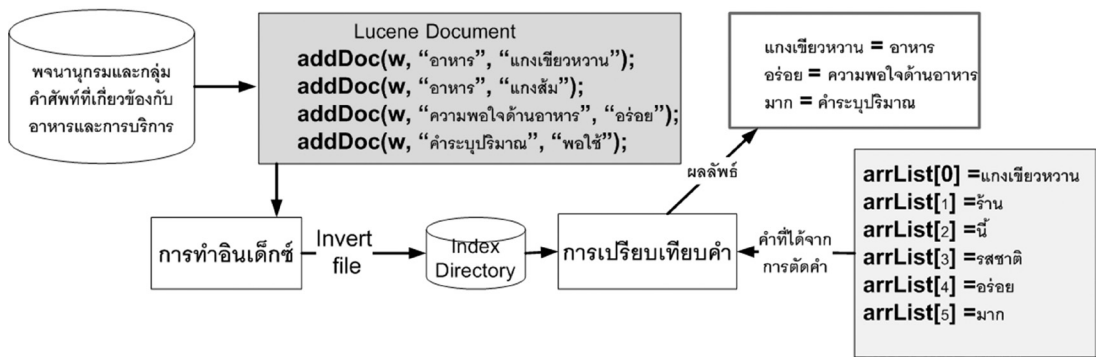
3. การสกัดคำ

ในขั้นตอนนี้จะเกี่ยวข้องกับการเปรียบเทียบคำที่ตัดได้ในประโยคกับคำศัพท์ในโดเมนร้านอาหารเพื่อทำการระบุว่ามีพจน์หรือคำที่ตัดได้นั้นเป็นคำประเภทใด จากนั้นจะทำการวิเคราะห์รูปแบบประโยคที่สมบูรณ์ (ตามหลักไวยากรณ์) เพื่อการนำไปใช้ในการคำนวณความคิดเห็นในขั้นตอนต่อไป

3.1 การเปรียบเทียบคำที่ตัดได้กับคำศัพท์ในโดเมนร้านอาหาร

สำหรับขั้นตอนนี้ก่อนอื่นผู้วิจัยได้นำพจนานุกรม LEXiTRON ที่อยู่ในรูปแบบ XML มาสกัดและจัดเก็บในฐานข้อมูล นอกจากนี้ยังรวบรวมคำนาม คำคุณศัพท์และคำวิเศษณ์ในโดเมนร้านอาหารและการบริการจากเว็บไซต์วีวอาหาร ซึ่งบางคำเป็นคำในภาษาพูดแบบและบางคำเป็นคำศัพท์แสดง คำศัพท์ที่รวบรวมมาได้นี้จะถูกจัดเก็บลงในตาราง คำศัพท์อาหารและบริการ ที่ประกอบไปด้วย 3 แอตทริบิวต์ คือ termId (ลำดับ), object (ชนิดของคำศัพท์แบ่งตามคุณลักษณะ: อาหาร/บริการ/บรรยากาศ/ความสะอาด/ความคุ้มค่า/ความพอใจ/ค่าระบุนปริมาณ) และ terminology (คำศัพท์) จากนั้นจะนำคำที่ตัดได้มาเปรียบเทียบกับคำศัพท์ในตาราง คำศัพท์อาหารและบริการ

การเปรียบเทียบจะใช้ Lucene API [15] ซึ่งเป็นเครื่องมือหลักในการสร้างอินเด็กซ์และสืบค้นข้อมูลโดยอาศัยโมเดลการสืบค้นแบบ Vector space ขั้นตอนการทำงานจะเริ่มจากการนำคำศัพท์และชนิดของคำที่ได้จากตารางคำศัพท์อาหารและบริการ มาสร้างเป็นอินเด็กซ์เพื่อใช้สำหรับการเปรียบเทียบ โดยระบบจะทำการรวบรวมคำที่อยู่ในอาร์เรย์ (ArrayList) มาเพื่อหาค่าความคล้ายคลึงกับคำศัพท์ที่เราได้ทำอินเด็กซ์ไว้ หากพบว่าคำนั้นมีความคล้ายคลึงกับเซตของอินเด็กซ์ในโดเรทอริกก็จะระบุว่าคำนั้นมีความสัมพันธ์กับชนิดของคำศัพท์ที่ได้อธิบายมาแล้วข้างต้น เช่นในภาพที่ 3 พบว่า arrList[0] เก็บคำว่า “แกงเขียวหวาน” จากนั้นจะนำคำว่า “แกงเขียวหวาน” ไปสืบค้นในอินเด็กซ์โดเรทอริกก็จะได้ว่าสอดคล้องกับคำศัพท์ชนิด “อาหาร” arrList[4] เก็บคำว่า “อร่อย” หลังจากเทียบกับอินเด็กซ์แล้วจะรู้ว่าเกี่ยวข้องกับ “ความพอใจ” และคำว่า “มาก” ในอาร์เรย์ลำดับต่อมาสามารถระบุได้ว่าเป็น “คำระบุปริมาณ” เป็นต้น เมื่อทำการรับคำจากอาร์เรย์ทั้งหมดแล้วก็จะนำคำที่ระบุชนิดได้ทั้งหมดมาเก็บในอาร์เรย์ผลลัพธ์ที่ชื่อ Extract เพื่อที่จะวิเคราะห์ว่ามีรูปประโยคที่สมบูรณ์พอสำหรับการวิเคราะห์คะแนนความคิดเห็นหรือไม่



ภาพที่ 3 การเปรียบเทียบคำที่ตัดได้กับคำศัพท์ใน โดเมนร้านอาหาร

3.2 การวิเคราะห์รูปแบบประโยคที่สมบูรณ์

ในการหารูปประโยคที่สมบูรณ์จะพิจารณาจากการเรียงตัวของคำในอาร์เรย์ Extract ของแต่ละประโยคความคิดเห็นตามไวยากรณ์แบบไม่พึ่งบริบท (Context Free Grammar) กล่าวคือจะพิจารณาหน้าที่คำในประโยค (S) โดยจะพิจารณาคำนาม (N) คำนามวลี (NP) กริยาวลี (VP) นิเสธ (NEG) คำคุณศัพท์ (ADJ) คำประมาณวิเศษณ์เพื่อบ่งบอกปริมาณ (ADVQ) ที่เกี่ยวข้องกับคุณลักษณะของร้านอาหารใน 5 ด้าน เรียงตามรูปแบบนิพจน์ปกติของไวยากรณ์ภาษาไทยที่กำหนดไว้ ยกตัวอย่างเช่น

$$S = NP + VP$$

$$NP = N|N + (ADJ) + (ADVQ) + (PP)|PRON$$

จากนั้นจะวิเคราะห์ข้อความความคิดเห็นของประโยค โดยพิจารณาจาก (1) รูปประโยคเป็นปฏิเสธหรือบอกเล่า (2) ความหมายของคำคุณศัพท์นั้นๆ เช่น รูปประโยค แกงเขียวหวาน/ อร่อยที่สุด เราจะพบว่า “แกงเขียวหวาน” เป็นคำนามที่เกี่ยวข้องกับความคิดเห็นด้านอาหาร “อร่อย” เป็นคำคุณศัพท์ที่เกี่ยวข้องกับรสชาติอาหารที่สามารถระบุความพึงพอใจ และคำว่า “ที่สุด” จะเป็นคำที่บ่งบอกปริมาณความพึงพอใจว่าอยู่ในระดับใด

4. การคำนวณคะแนนความคิดเห็น

การคำนวณคะแนนความพึงพอใจจากความคิดเห็นของผู้ใช้จะอาศัยค่าน้ำหนักของคำระบุปริมาณ โดยค่าน้ำหนักนี้ได้มาจากการสำรวจเว็บไซต์รีวิวร้านอาหารต่างๆ ที่ให้ผู้ใช้แสดงความคิดเห็นและมีค่าเหล่านั้นปรากฏอยู่จากนั้นคำนวณออกมาเป็นค่าระดับน้ำหนัก เช่น คำว่า “ที่สุด” มีค่าน้ำหนักเท่ากับ 2 คำว่า “มาก” มีค่าน้ำหนักเท่ากับ 1.6 หากไม่ปรากฏคำระบุปริมาณในประโยค ให้ค่าน้ำหนักเท่ากับ 1.6 เป็นต้น โดยค่าคะแนนของประโยคความคิดเห็นจะเป็นบวกหรือลบขึ้นอยู่กับความพึงพอใจและน้ำหนักของคำระบุปริมาณ เช่นถ้าผู้ใช้เขียนว่า “ซอสนสปรกที่สุด” ค่าคะแนนด้านความสะอาดก็จะ เป็น - 2 (สปรก ความพึงพอใจติดลบ, ที่สุด ค่าน้ำหนักของคำระบุปริมาณ = 2) ตัวอย่างการให้คะแนนโดยอาศัยน้ำหนักของคำในประโยคแสดงในตารางที่ 1

ตารางที่ 1 ตารางข้อมูล term extraction ที่เก็บข้อมูลค่าคะแนนความคิดเห็นจากแต่ละประโยคของผู้ใช้

Id	ext_restaurant	Ext_token	Ext_object	opinion	Ext_category	volume	score
1	ร้านบ้านฝนตก	สเด็ก	อาหาร	อร่อย	ความคิดเห็นเชิงบวก	-	1.6
2	ร้านบ้านฝนตก	พนักงาน	บริการ	ห่วย	ความคิดเห็นเชิงลบ	มาก	-1.8
3	ครัวผู้ใหญ่บ้าน	จาน	ความสะอาด	สปรก	ความคิดเห็นเชิงลบ	ที่สุด	-2

ในการคำนวณค่าความคิดเห็นต่อร้านอาหาร เริ่มจากการหาค่าเฉลี่ยความคิดเห็นของผู้ใช้แต่ละคนต่อความพึงพอใจในแต่ละด้าน ตามสมการ (a)

$$y = \frac{(x_1 + x_2 + x_3 + \dots + x_n)}{n} \quad (a)$$

กำหนดให้ y แทน ค่าเฉลี่ยความคิดเห็นของผู้ใช้แต่ละคน

x แทน คะแนนความคิดเห็นย่อยที่เกิดขึ้นในแต่ละครั้ง

n แทน จำนวนความคิดเห็นที่เกิดขึ้นในแต่ละด้านจากผู้ใช้

ยกตัวอย่างการคำนวณ เช่น นายสมชาย ได้ให้ความเห็นด้านอาหารของร้านบ้านฝนตกไว้ว่า “สเด็กเนื้ออร่อยมาก / เนื้อนุ่มมาก / แต่แกงส้มรสชาติจืดจาง” เราจะพบว่าความคิดเห็นดังกล่าวประกอบด้วย 3 ประโยคย่อย : (1) สเด็กเนื้ออร่อยมาก มีคะแนนเท่ากับ 1.8 (2) เนื้อนุ่มมาก มีคะแนนเท่ากับ 1.8 และ (3) แกงส้มรสชาติจืดจาง มีคะแนนเท่ากับ 1.6 ดังนั้นคะแนนเฉลี่ยของสมชายที่มีต่อร้านอาหารบ้านฝนตกในด้านอาหารมีค่าเท่ากับ $(1.8+1.8+1.6)/3 = 1.73$ จากนั้นระบบจะนำความคิดเห็นด้านอาหารของร้านบ้านฝนตกจากผู้ใช้ทุกคนมาหาค่าเฉลี่ยตามสมการ (b)

$$z = \frac{(y_1 + y_2 + y_3 + \dots + y_m)}{2m} \quad (b)$$

กำหนดให้ z แทน คะแนนความพึงพอใจเฉลี่ยจากผู้ให้บริการในด้านนั้นๆ

m แทน จำนวนความคิดเห็นจากผู้ใช้งานในด้านนั้นๆ

5. การแสดงผลความพึงพอใจ

หลังจากคำนวณคะแนนความคิดเห็นแต่ละคุณลักษณะของแต่ละร้านแล้ว ขั้นตอนต่อมาที่จะทำการแสดงผลคะแนนความพึงพอใจแต่ละด้านเป็นรูปดาวแบ่งตามช่วงคะแนน ทั้งนี้เพื่อให้ผู้ใช้งานที่สืบค้นข้อมูลร้านอาหารเข้าใจภาพรวมของคุณลักษณะแต่ละด้าน ได้ง่ายขึ้น

การทดสอบและอภิปรายผล

ผู้วิจัยได้ทำการประเมินผลเพื่อศึกษาว่าระบบสามารถสรุปความคิดเห็นเกี่ยวกับร้านอาหารได้สอดคล้องกับการสรุปโดยผู้ใช้งานน้อยเพียงใด ในการประเมินจะมีผู้ร่วมประเมินจำนวน 20 คน ซึ่งเป็นนิสิตระดับปริญญาตรีจำนวน 13 คน และนิสิตระดับบัณฑิตศึกษา 7 คน นิสิตทั้งหมดศึกษาในสาขาวิทยาการคอมพิวเตอร์และสาขาเทคโนโลยีสารสนเทศ โดยผู้ร่วมประเมินมีประสบการณ์ใช้เว็บไซต์รีวิร้านอาหารและเคยเขียนแสดงความคิดเห็นเกี่ยวกับร้านอาหารที่พวกเขาเคยไปชิม สำหรับขั้นตอนการประเมินนั้นผู้ประเมินจะได้รับรายการความคิดเห็นเกี่ยวกับร้านอาหารจำนวน 5 ร้าน ร้านละ 100 ความคิดเห็น ซึ่งผู้วิจัยได้นำความคิดเห็นเหล่านี้มาจากเว็บไซต์ลงในผู้ร่วมประเมินอ่านกลุ่มความคิดเห็นทดสอบและทำการให้ดาวในแต่ละด้านตามที่พวกเขาคิดว่าสอดคล้องกับเนื้อหาความคิดเห็น รายการความคิดเห็นชุดเดียวกันจะนำไปเป็นอินพุตของระบบ จากนั้นผู้ร่วมประเมินจะทำการเปรียบเทียบจำนวนดาวจากการสรุปของพวกเขากับจำนวนดาวที่ได้จากการสรุปความคิดเห็นจากระบบก่อนจะประเมินความพึงพอใจด้านความสอดคล้องของการสรุปผ่านแบบสอบถาม โดยคำถามแต่ละข้อจะมีระดับความพึงพอใจ 1-5 จากนั้นผู้วิจัยจะนำแบบสอบถามทั้งหมดมาหาค่าเฉลี่ยของคำถาม โดยใช้เกณฑ์ในการวัดค่าเฉลี่ยความพึงพอใจดังแสดงในตารางที่ 2

ตารางที่ 2 เกณฑ์ในการวัดระดับความพึงพอใจด้านความสอดคล้องของการสรุปความคิดเห็น

เชิงปริมาณ	เชิงคุณภาพ
4.51-5.00	ดีมาก
3.51-4.50	ดี
2.51-3.50	พอใช้
1.51-2.50	ปรับปรุง
1.00-1.50	ไม่เหมาะสม

ผลการประเมินแสดงในตารางที่ 3 พบว่าค่าความพึงพอใจด้านความสอดคล้องในการสรุปความคิดเห็นสำหรับคุณลักษณะด้านบริการและอาหารอยู่ในระดับดี ส่วนด้านอื่นๆ ที่เหลืออยู่ในระดับพอใช้

ตารางที่ 3 ผลการประเมินความความสอดคล้องในการสรุปผลระหว่างระบบกับผู้ใช้

รายการประเมิน	ระดับความพึงพอใจ	
	เชิงปริมาณ	เชิงคุณภาพ
ความสอดคล้องในการสรุปผลด้านการบริการ	3.68	ดี
ความสอดคล้องในการสรุปผลด้านบรรยากาศ	3.28	พอใช้
ความสอดคล้องในการสรุปผลด้านความสะดวก	2.87	พอใช้
ความสอดคล้องในการสรุปผลด้านราคา	3.04	พอใช้
ความสอดคล้องในการสรุปผลด้านอาหาร	3.52	ดี
ค่าเฉลี่ยรวม	3.27	พอใช้

สรุปผลการวิจัย

งานวิจัยนี้เสนอระบบวิเคราะห์และสรุปความคิดเห็นเกี่ยวกับร้านอาหารจากเว็บไซต์วีว โดยอัตโนมัติที่ใช้เทคนิคการตัดคำในประโยคโดยอาศัยฐานคำศัพท์ การวิเคราะห์ประเภทคำและรูปประโยค เพื่อหาความหมายเชิงบวกหรือเชิงลบของประโยคและนำมาคำนวณเป็นค่าคะแนนความพึงพอใจสำหรับปัจจัยด้านการบริการต่างๆ ในรูปแบบกราฟิก ผลการประเมินความพอใจด้านความสอดคล้องในการสรุปความคิดเห็นระหว่างระบบกับผู้ใช้ อยู่ในระดับพอใช้ ซึ่งเป็นผลมาจากข้อจำกัดที่เกี่ยวข้องกับการประมวลผลภาษาธรรมชาติในหลายประเด็นเช่น (1) ผู้ใช้แสดงความคิดเห็น โดยใช้รูปประโยคที่หลากหลายและบางครั้งไม่ถูกหลักไวยากรณ์ เช่น “เบคอนทอดมากรอบกำลังดีและมีรสเค็มนิดๆจากเบคอน ลงตัวมากๆกับความหวานของเห็ด” จะเห็นว่าประโยคแรกเป็นภาษาพูด และประโยคที่สองมีการเรียงชนิดของคำไม่ถูกต้องตามหลักไวยากรณ์ (2) ความคิดเห็นบางประโยคเป็นแบบเชิงซ้อนและแบบอุปมาอุปมัยที่ไม่สามารถตีความได้โดยตรง เช่น “กินส้มตำร้านนี้แทบจะฟันไฟได้” เมื่อตีความแล้วจะหมายถึงเผ็ดมาก เป็นต้น ซึ่งข้อจำกัดเหล่านี้เป็นงานในอนาคตที่ผู้วิจัยต้องศึกษา ซึ่งได้แก่ การขยายขีดความสามารถของระบบให้สามารถวิเคราะห์ประโยคเชิงซ้อนได้ และการศึกษาเปรียบเทียบเทคนิคการวิเคราะห์รูปประโยคแบบต่างๆ เพื่อสนับสนุนการสกัดคำที่ถูกต้องและตรงความหมายของผู้ใช้มากขึ้น

เอกสารอ้างอิง

[1] Feldman, R . (2013). “Techniques and Applications for Sentiment Analysis”, **Communications of the ACM**. 56(4), 82-89.

[2] Hussein, D.-M.E.D.M. (2016). “A Survey on Sentiment Analysis Challenge”, **Journal of King Saud University Engineering Science** [Online]. Available: <http://dx.doi.org/10.1016/j.jksues.2016.04.002>.

[3] Phawattanakul, K and Luenam, P. (2556). “Opinion Mining from Online Social Networks”, **Modern Management Journal**. 11(2), 11-20.

[4] Xia, R., Xu, F., Yu, J., Qi, Y and Cambria, E. (2016). “Polarity Shift Detection, Elimination and Ensemble: A three-stage model for document-level sentiment analysis”, **Information Processing and Management**. 52, 36-45.

- [5] Turney, P. D. (2002). “Thumbs Up or Thumbs Down ? Semantic Orientation Applied to Unsupervised Classification of Reviews”, In **Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)**. Association for Computational Linguistics. Pennsylvania: Association for Computational Linguistics Stroudsburg. 417-424.
- [6] Sangkeetrakarn. (n.d). **FANALYTICS: Sentiment Analysis of Messages on Facebook Wall** [Online]. Available: [http:// conan.in.th](http://conan.in.th) [Accessed 12th January 2015].
- [7] Speech and Audio Technology Laboratory (SPT) National Electronic and Computer Center (NECTEC). (n.d) **S-Sense** [Online]. Available: <http://www.ssense.in.th/> [Accessed 12th February 2016].
- [8] Gavilanes, F, M., Lopez, A, T. and Martinez, J, J. (2016). “Unsupervised Method for Sentiment Analysis in Online Texts”, **Expert Systems with Applications**. 58, 57-75.
- [9] Prombut, N. and Temtanapat, Y. (2008). “Mining Opinion in Product Reviews : A Case Study of Mobile Phone Reviews”, In **The 12th National Computer Science and Engineering Conference (NCSEC2008)**. 257-263.
- [10] Wonsri, W. (2010). **Opinion Mining System on Hotel Reviews (Research report)**. Bangkok: National Science and Technology development Agency, National Electronic and Computer Center.
- [11] Pang, B. and Lee, L. (2008). “Opinion Mining and Sentiment Analysis”, **Journal of the ACM**. 2, 1-135.
- [12] Wongsin, C., Srikarnjanaperd, N. and Polpinit, J. (2015). “Automatic Feeling Analysis from Opinion Text”, In **The 11th National Conference on Computing and Information Technology**. 1-6.
- [13] Haruechaiyasak, C. (2006). **LongLexto**. Tokenizing Thai texts using longest matching approach.
- [14] LEXiTRON. (2009). **LEXiTRON** [Online]. Available: [http:// http://lexitron.nectec.or.th/2009_1/](http://lexitron.nectec.or.th/2009_1/) [Accessed 20th November 2015].
- [15] Lucene. (2010). **Lucene** [Online]. Available: <http://lucene.apache.org/java/docs/index.html> [Accessed 25th November 2015].