

การสกัดหัวข้อของรีวิรร้านอาหารในกรุงเทพบนเว็บไซต์

ASPECT TERM EXTRACTION OF RESTAURANT REVIEW IN  
BANGKOK ON WEBSITE



การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์  
ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง  
คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2566

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้ KMUTL-2023-SC-M-017-030 อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ASPECT TERM EXTRACTION OF RESTAURANT REVIEW IN  
BANGKOK ON WEBSITE



AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE  
IN DATA SCIENCE AND ANALYTICS  
KMITL DIGITAL ANALYTICS AND INTELLIGENCE CENTER SCHOOL OF SCIENCE  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG  
2023

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

KMITL-2023-SC-M-017-030



COPYRIGHT 2023

SCHOOL OF SCIENCE

**KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อการค้นคว้าอิสระ	การสกัดหัวข้อของรีวิวร้านอาหารในกรุงเทพบนเว็บไซต์
ชื่อนักศึกษา	ศุภกร โกมลหทัย
รหัสประจำตัว	6460510964
ปริญญา	วิทยาศาสตรมหาบัณฑิต (วิทยาการข้อมูลและการวิเคราะห์)
พ.ศ.	2566
อาจารย์ที่ปรึกษาการค้นคว้าอิสระ	รองศาสตราจารย์ ดร.ละออ บุญเกษม

### บทคัดย่อ

ในปัจจุบันบทวิจารณ์ร้านอาหารบนโลกออนไลน์เกิดขึ้นอย่างมากมาย เพื่อให้ลูกค้าผู้ใช้บริการร้านอาหารมีส่วนร่วมในการแสดงความคิดเห็น บอกเล่าความประทับใจ นำเสนอข้อปรับปรุงในการใช้บริการร้านอาหาร ทำให้ผู้ประกอบการร้านอาหารตอบสนองการให้บริการต่อลูกค้าได้ดียิ่งขึ้น ลูกค้าสามารถตัดสินใจเลือกใช้บริการร้านอาหารได้ง่ายยิ่งขึ้นจากบทวิจารณ์ร้านอาหารออนไลน์ ซึ่งการสกัดคุณลักษณะของข้อความ (Aspect) ช่วยให้ลูกค้าเข้าใจบริบทได้อย่างรวดเร็วที่เกิดขึ้นในบทวิจารณ์ทั้งหมด ลดเวลาในการวิเคราะห์จุดเด่น จุดด้อยต่างๆของร้านค้า การวิเคราะห์คุณลักษณะประกอบด้วย 3 วิธีการดังนี้ 1) วิธีการสกัดคุณลักษณะเชิงความถี่ด้วยการคำนวณ TF-IDF ช่วยสกัดบริบทต่างๆที่เกิดขึ้นบ่อยจำนวนมากได้ 2) วิธีการสกัดคุณลักษณะเชิงโครงสร้างด้วยกฎไวยากรณ์ ช่วยให้สามารถสกัดคุณลักษณะที่เกิดขึ้นตามโครงสร้างของประโยคได้ 3) การสกัดคุณลักษณะเชิงความถี่ร่วมกับคุณลักษณะเชิงโครงสร้าง เป็นการรวมข้อดีของการสกัดคุณลักษณะทั้งสองชนิดเข้าด้วยกัน ทำให้คุณลักษณะที่ได้จากบทวิจารณ์มีความสมบูรณ์มากยิ่งขึ้น ทั้งนี้ผู้วิจัยได้ประยุกต์คลังคำศัพท์คำพ้องภาษาไทยในการจำแนกกลุ่มคำคุณลักษณะที่เกิดขึ้น เพื่อให้เกิดคุณลักษณะที่สอดคล้องกับคำในภาษาไทยมากยิ่งขึ้น การสกัดคุณลักษณะเชิงความถี่ร่วมกับโครงสร้างจะให้ผลของการสกัดคุณลักษณะที่มีความหลากหลายตามธรรมชาติภาษาของข้อมูล เนื่องจากมีการสกัดโครงสร้างตามกฎไวยากรณ์ และมีการคัดกรองคำที่มีความน่าหนักความถี่ และค่าความถี่ของประโยคที่พบคุณลักษณะตามที่กำหนด ผู้เชี่ยวชาญสามารถกำหนดผลเฉลยคุณลักษณะที่ใช้เป็นข้อสรุปบทวิจารณ์ร้านอาหารโดยรวมได้ในอนาคต เพื่อใช้สำหรับการคำนวณประสิทธิภาพเชิงตัวเลข

**คำสำคัญ:** บทวิจารณ์, คุณลักษณะ, ร้านอาหาร, ความถี่, กฎไวยากรณ์

<b>Independent Study Title</b>	Aspect term extraction of restaurant review in Bangkok on website
<b>Student Name</b>	Supakorn Komolhatai
<b>Student ID</b>	6460510964
<b>Degree</b>	Master of Science (Data Science and Analytics) KMITL-Digital Analytics and Intelligence Center
<b>Year</b>	2023
<b>Independent Study Advisor</b>	Assoc. Prof. Dr. Laor Boongasame

### Abstract

There are numerous online restaurant reviews where customers can participate by expressing their opinions, sharing their experiences, and suggesting improvements for better restaurant service. These reviews enable restaurant owners to better respond to customer needs and make it easier for customers to decide which restaurant to choose. The analysis of review texts involves identifying aspects to quickly understand the context within all the reviews. Aspect extraction consists of three methods: 1) Frequency-based aspect extraction using TF-IDF calculation helps extract frequently occurring contexts effectively. 2) Structure-based aspect extraction using grammatical rules allows the extraction of aspects based on sentence structures. 3) Combined frequency-based and structure-based aspect extraction, which combines the advantages of both methods to obtain comprehensive aspect analysis from the reviews. Researchers have applied Thai language synonym dictionaries to classify aspect-related words more accurately. The combined frequency-based and structure-based approach provides diverse aspect extraction results based on the natural characteristics of the Thai language data. It involves extracting structures based on grammatical rules and filtering keywords based on their frequency and the frequency values of the predefined aspects. Experts can determine the final aspect results for overall restaurant reviews in the future, which can be used for performance evaluation.

**Keywords:** Review, Aspect Term, Restaurant, Term Frequency - Inverse Document Frequency (TF-IDF), Rule-based

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## กิตติกรรมประกาศ

งานวิจัยฉบับนี้ข้าพเจ้าขอกราบขอบพระคุณ รศ.ดร.ละออ บุญเกษม อาจารย์ที่ปรึกษา ให้ความรู้ และคำแนะนำในการปรับปรุงแก้ไขงานวิจัยให้สำเร็จลุล่วงไปได้ด้วยดี มีประสิทธิภาพ ตลอดจนให้คำแนะนำในเรื่องต่างๆ ซึ่งผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง

สุดท้ายนี้ขอกราบขอบพระคุณบิดามารดา ตลอดจนญาติสนิทมิตรสหายที่คอยให้ กำลังใจ และคำแนะนำในการสร้างสรรค์งานวิจัยชิ้นนี้ให้ได้เริ่มต้นขึ้น



นาย ศุภกร โกมลหทัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	ฉ
สารบัญรูป	ช
<b>บทที่ 1 บทนำ</b>	<b>1</b>
1.1 ความเป็นมาและความสำคัญของงานวิจัย	1
1.2 วัตถุประสงค์ของงานวิจัย	2
1.3 ขอบเขตของงานวิจัย	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	2
<b>บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง</b>	<b>3</b>
2.1 ทฤษฎีที่เกี่ยวข้อง	3
2.1.1 การเรียนรู้แบบไม่มีผู้สอน	3
2.1.2 การประมวลผลทางภาษารธรรมชาติ	3
2.1.3 วิเคราะห์คุณลักษณะจากการประมวลผลทางภาษา	4
2.1.4 คำพ้องความหมาย	6
2.1.5 แบบจำลองภาษา N-Gram	6
2.1.6 วิธีการเชิงความถี่	7
2.1.7 วิธีการใช้กฎฐาน	8
2.2 งานวิจัยที่เกี่ยวข้อง	8
2.2.1 กลุ่มงานวิจัยเกี่ยวกับปัจจัยในการวิเคราะห์ร้านอาหาร	11
2.2.2 กลุ่มงานวิจัยการวิเคราะห์ข้อความทางภาษาศาสตร์	13
2.2.3 กลุ่มงานวิจัยการวิเคราะห์ข้อความแบบ Aspect Term Sentiment	13
2.2.4 กลุ่มงานวิจัยการวิเคราะห์ข้อความแบบ Aspect Category Sentiment	14
<b>บทที่ 3 วิธีการดำเนินงานวิจัย</b>	<b>17</b>
3.1 การจัดเก็บรายชื่อร้านอาหารบนแพลตฟอร์มออนไลน์	19
3.2 การจัดเก็บบทวิจารณ์ร้านอาหารแต่ละร้าน	19
3.3 การจัดเตรียมคลังข้อมูลคำพ้องความหมาย (WordNet)	20
3.4 การจัดเตรียมข้อมูลบทวิจารณ์ร้านอาหารออนไลน์	20

เอกสารนี้เป็นเอกสารที่ 3.4.1 การปรับข้อความเป็นตัวพิมพ์เล็ก (lower case) ชาติให้นำไปใช้ประโยชน์ด้าน 20 คำ  
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.2 การลบอักขระพิเศษ และอีโมจิ	21
3.4.3 การกำหนดค่าโมเดลการตัดคำ (Tokenizer)	21
3.4.4 การติดแท็ก Part-Of-Speech (POS)	21
3.4.5 การขจัดแท็กชนิดคำที่ไม่จำเป็น	22
3.5 การสกัดคุณลักษณะของผลิตภัณฑ์เชิงความถี่	22
3.5.1 การหาค่า Term frequency - Inverse Document Frequency	22
3.5.2 การค้นหาคำแทนคุณลักษณะโดยใช้เซตคำพ้องความหมาย	27
3.6 การสกัดคุณลักษณะของผลิตภัณฑ์เชิงโครงสร้าง	28
3.7 การสกัดคุณลักษณะของผลิตภัณฑ์เชิงความถี่ร่วมกับโครงสร้าง	29
<b>บทที่ 4 ผลการวิจัยและการอภิปรายผล</b>	31
4.1 การสกัดคุณลักษณะของผลิตภัณฑ์เชิงความถี่	31
4.2 การสกัดคุณลักษณะของผลิตภัณฑ์เชิงโครงสร้าง	34
4.3 การสกัดคุณลักษณะของผลิตภัณฑ์เชิงความถี่ร่วมกับโครงสร้าง	38
<b>บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ</b>	42
เอกสารอ้างอิง	44
ประวัติผู้เขียน	46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญตาราง

ตารางที่	หน้า
2.1 รายการงานวิจัยจำแนกตามหมวดหมู่	10
3.1 ตัวอย่างการจัดเก็บรายชื่อร้านอาหารบนแพลตฟอร์มออนไลน์	19
3.2 ตัวอย่างการจัดเก็บบทวิจารณ์ร้านอาหารแต่ละร้าน	19
3.3 ตัวอย่างการจัดเก็บคลังข้อมูลคำพ้องความหมาย	20
3.4 การคำนวณค่า TF value บทวิจารณ์ที่ 1	22
3.5 การคำนวณค่า TF value บทวิจารณ์ที่ 2	23
3.6 การคำนวณค่า ICF value บทวิจารณ์ที่ 1	24
3.7 การคำนวณค่า ICF value บทวิจารณ์ที่ 2	25
3.8 การคำนวณค่า TF-ICF value บทวิจารณ์ที่ 1	25
3.9 การคำนวณค่า TF-ICF value บทวิจารณ์ที่ 2	26
3.10 การคัดเลือกค่าแทนคุณลักษณะจาก TF-ICF ของแต่ละคำ	27
3.11 กฎไวยากรณ์การสกัดคุณลักษณะผลิตภัณฑ์เชิงโครงสร้าง	29
3.12 การสกัดคุณลักษณะเชิงความถี่ร่วมกับโครงสร้าง	29
4.1 จำนวนคุณลักษณะเชิงความถี่ในแต่ละกลุ่มร้านอาหาร	31
4.2 การจัดกลุ่มคุณลักษณะเชิงความถี่ร้านอาหารญี่ปุ่น	32
4.3 การจัดกลุ่มคุณลักษณะเชิงความถี่ร้านอาหารบุฟเฟ่ต์	33
4.4 กฎไวยากรณ์	35
4.5 จำนวนคุณลักษณะเชิงโครงสร้างในแต่ละกลุ่มร้านอาหาร	36
4.6 การจัดกลุ่มคุณลักษณะเชิงความถี่โครงสร้างร้านอาหารญี่ปุ่น	36
4.7 การจัดกลุ่มคุณลักษณะเชิงความถี่โครงสร้างร้านอาหารบุฟเฟ่ต์	37
4.8 จำนวนคุณลักษณะเชิงความถี่ร่วมกับโครงสร้างในแต่ละกลุ่มร้าน	38
4.9 การจัดกลุ่มคุณลักษณะเชิงความถี่ร่วมกับโครงสร้างร้านอาหารญี่ปุ่น	39
4.10 การจัดกลุ่มคุณลักษณะเชิงความถี่ร่วมกับโครงสร้างร้านอาหารบุฟเฟ่ต์	40

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญรูป

รูปที่	หน้า
2.1 ตัวอย่างการสกัดคุณลักษณะจากความคิดเห็น	5
2.2 ตัวอย่างการสกัดความรู้สึกจากคุณลักษณะ	5
2.3 ตัวอย่างการจัดประเภทคุณลักษณะ	6
2.4 ตัวอย่างการวิเคราะห์ความรู้สึกจากประเภทคุณลักษณะ	6
2.5 ตัวอย่างการจำแนกคุณลักษณะของข้อความในรูปแบบ 1-gram (unigram)	7
2.6 Restaurant Quality Hypothesis	11
2.7 Results of regression analysis (1)	12
2.8 Results of regression analysis (2)	12
2.9 Accuracy of Hindi Dependency Parser	13
2.10 Results of restaurant domain analysis	16
2.11 Results of hotel domain analysis	16
3.1 แผนภาพการทำงานของอัลกอริทึมสกัดคำคุณลักษณะ	18
3.2 ตัวอย่างการปรับข้อความเป็นตัวพิมพ์เล็ก	20
3.3 ตัวอย่างการจัดอักขระพิเศษ และอีโมจิ	21
3.4 ตัวอย่างการใช้งานโมเดลการตัดคำ	21
3.5 ตัวอย่างการแท็ก POS	21
4.1 ตัวอย่างประโยคที่พบคำคุณลักษณะเชิงความถี่ “พัฒนา”	34
4.2 ตัวอย่างประโยคที่พบคำคุณลักษณะเชิงความถี่ “ปรับปรุง”	34
4.3 ตัวอย่างประโยคที่พบคำคุณลักษณะเชิงโครงสร้าง “การพัฒนา”	38
4.4 ตัวอย่างประโยคที่พบคำคุณลักษณะเชิงโครงสร้าง “ส่วนลด”	38
4.5 ตัวอย่างประโยคที่พบคำคุณลักษณะเชิงความถี่ร่วมโครงสร้าง “ภาษี”	40
4.6 ตัวอย่างประโยคที่พบคำคุณลักษณะเชิงความถี่ร่วมโครงสร้าง “ส่วนลด”	40

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันนี้การใช้เทคโนโลยีออนไลน์ได้เข้ามามีบทบาทต่อร้านอาหาร ทำให้วิถีชีวิตของคนเราได้เปลี่ยนแปลงไปอย่างมากมาย จากข้อมูลของ Wongnai พบว่าในปี 2018-2019 ธุรกิจร้านอาหารมีอัตราการเติบโตที่สูงขึ้น 40% และพบว่า 63% ของลูกค้านิยมการสั่งอาหารแบบ delivery นั้นทำให้ธุรกิจร้านอาหารมีช่องทางในการหารายได้ และมีโอกาสเจาะกลุ่มลูกค้าได้หลากหลายมากยิ่งขึ้น ทำให้ร้านค้าสามารถทราบถึงความต้องการลูกค้าได้มากขึ้น จากคำวิจารณ์ผ่านช่องทางออนไลน์ นอกเหนือจากข้อมูลที่ได้รับปากต่อปากจากร้านค้า สิ่งต่างๆเหล่านี้ทำให้ร้านอาหารสามารถพัฒนาความต้องการของลูกค้าได้ดียิ่งขึ้น ซึ่งคำวิจารณ์ และคะแนนรีวิวเข้ามาเป็นส่วนหนึ่งในการตัดสินใจเข้าใช้บริการได้เป็นอย่างดี ทำให้ร้านค้ามีความน่าสนใจ อาจเป็นตัวเลือกแรกในการตัดสินใจ

ความคิดเห็นต่างๆที่ร้านอาหารได้รับมานั้นสามารถนำมาวิเคราะห์ได้หลากหลาย ทั้งในด้านการวิเคราะห์เชิงบวกเชิงลบ และการวิเคราะห์มุมมองที่มีต่อร้านค้า ในอดีตการรับคำวิจารณ์อาจต้องใช้เวลาในการเก็บข้อมูล แต่ในปัจจุบันเทคโนโลยีออนไลน์ทำให้ข้อมูลเพิ่มขึ้นอย่างรวดเร็ว คำวิจารณ์ที่ได้มาจึงเป็นข้อมูลที่ใหม่ และรวดเร็วเสมอ เราจึงสามารถนำข้อมูลมาวิเคราะห์ได้มากยิ่งขึ้น ในปัจจุบันนี้มีการวิเคราะห์เชิงบวกเชิงลบกันอย่างแพร่หลาย ซึ่งนั่นอาจทำให้เห็นถึงมุมมองด้านต่างๆของแต่ละร้านไม่ได้มากเท่าที่ควร ผู้เขียนจึงเห็นว่าเราสามารถเพิ่มเติมความสามารถเหล่านี้ นอกเหนือจากเชิงบวกเชิงลบได้ เพราะในปัจจุบันนี้มีเทคโนโลยีการวิเคราะห์ภาษาที่ทันสมัยเพิ่มขึ้น ทำให้เราสามารถเข้าใจ และวิเคราะห์ร้านอาหารในเชิงลึกได้มากยิ่งขึ้น

เราสามารถพบเห็นงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์มุมมองที่มีต่อข้อความ ซึ่งในปัจจุบันนี้ งานวิจัยการประยุกต์วิเคราะห์ทางภาษาไทยอาจพบน้อย โดยส่วนใหญ่แล้วอาจพบได้ในภาษาอังกฤษโดยส่วนใหญ่ ไม่ว่าจะเป็นตัวอย่างการวิเคราะห์มุมมองที่มีต่อข้อความบนโลกโซเชียลเน็ตเวิร์คโดยใช้โมเดลทางภาษาที่มีการเทรนแล้ว โดยจะค้นหาข้อความที่เป็นเอนทิตีอาหารที่ทำจากพืช และค้นหาข้อความที่เกี่ยวข้องกับเอนทิตี เพื่อวิเคราะห์ออกมาเป็นกลุ่มเลเบลข้อความที่กำหนดไว้

สำหรับในงานวิจัยนี้ผู้เขียนสามารถนำไอเดียพัฒนาเพิ่มเติมได้จากงานวิจัยที่เกี่ยวข้อง โดยนำมาประยุกต์กับโมเดลทางภาษาไทยที่มีการเทรนแล้ว โดยการหาความสัมพันธ์ของคำ และนำมาวิเคราะห์มุมมองเพิ่มเติม ในอนาคตถ้าหากโมเดลทางภาษาไทยมีการพัฒนามากยิ่งขึ้น ก็อาจทำให้เราได้ประสิทธิภาพการทำงานที่ดียิ่งขึ้น และในปัจจุบันนี้การจำแนกจุดเด่นของร้านอาหารบนระบบเว็บไซต์รวบรวมร้านอาหารออนไลน์ ใช้การเก็บรวบรวมจากจำนวนการคำค้นหาที่มากที่สุด ผู้เขียนจึงเห็นว่าในอนาคตสามารถนำการวิเคราะห์มุมมองร้านอาหาร พัฒนาต่อยอดไปเป็นระบบการแนะนำร้านอาหารที่มีมุมมองตรงกับผู้บริโภคได้มากยิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 1.2 วัตถุประสงค์ของงานวิจัย

- 1) เพื่อพัฒนาอัลกอริทึมในการสกัดข้อความรีวิวในเชิงคุณลักษณะ
- 2) เพื่อศึกษาการใช้คำศัพท์คำพ้องความหมาย ในการจัดกลุ่มคุณลักษณะเชิงความถี่ และเชิงโครงสร้างของคุณลักษณะที่เกิดขึ้น

## 1.3 ขอบเขตของงานวิจัย

การวิจัยครั้งนี้นำข้อมูลการวิเคราะห์จากแพลตฟอร์มรีวิวร้านอาหารยอดนิยมในกรุงเทพมหานคร ที่อยู่ระหว่างปีพุทธศักราช 2561 จำนวน 2,700 รายการ ซึ่งเป็นรายการความคิดเห็นที่ไม่มีอักขระพิเศษ และรูปภาพ

## 1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ทำให้ทราบถึงคุณลักษณะของลูกค้าที่มีต่อร้านอาหารในกรุงเทพมหานครในเชิงแอชเพ็ค
- 2) ผลการศึกษาที่ได้สามารถนำไปเป็นข้อมูลเบื้องต้นในการตัดสินใจเลือกทานร้านอาหารได้
- 3) ทำให้ทราบถึงข้อจำกัดของการใช้งานอัลกอริทึมการสกัดคุณลักษณะ เพื่อนำไปประยุกต์ใช้กับการวิเคราะห์ความคิดเห็นข้อมูลกลุ่มธุรกิจอื่นได้ในอนาคต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 2

# ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

จากการศึกษาแนวคิด และทฤษฎีต่าง ๆ ที่ผู้เขียนได้รวบรวม เพื่อใช้ในงานวิจัยนี้ ประกอบไปด้วยทฤษฎีที่เกี่ยวข้องต่อไปนี้ การเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning), การประมวลผลทางภาษาธรรมชาติ, การวิเคราะห์คุณลักษณะความคิดเห็นจากการประมวลผลทางภาษา, คำพ้องความหมาย (Synonyms), วิธีการเชิงความถี่ (Frequency-based Approach), วิธีการใช้กฎฐาน (Rule-based Approach)

### 2.1 ทฤษฎีที่เกี่ยวข้อง

#### 2.1.1 การเรียนรู้แบบไม่มีผู้สอน

เป็นการเรียนรู้จากรูปแบบข้อมูลที่มีอยู่ สำหรับข้อมูลที่ไม่ระบุประเภท (Unlabeled) หรือมีคุณลักษณะของข้อมูลที่ไม่สามารถระบุได้ การเรียนรู้แบบไม่มีผู้สอนใช้ในการวิเคราะห์ข้อมูล โดยไม่พิจารณาที่ทำนายเป็นหลัก [2] โดยมากแล้วตัวอย่างการประยุกต์ใช้มีรูปแบบดังนี้ การสกัดคุณสมบัติข้อมูล (Feature extraction) การค้นหารูปแบบข้อมูลที่สำคัญ การจับคู่ข้อมูลระหว่างกัน การตรวจจับความผิดปกติ (Anomaly detection) การแบ่งกลุ่มข้อมูล (Clustering) เป็นเทคนิคที่นิยมใช้ในการแบ่งกลุ่มข้อมูล ประเภทของการแบ่งกลุ่มข้อมูล ได้แก่ เทคนิคการแบ่งกลุ่มแบบมีลำดับชั้น (Hierarchical clustering) เทคนิคการแบ่งกลุ่มแบบไม่มีลำดับชั้น (Non-hierarchical clustering) เป็นต้น การประมาณค่าความหนาแน่น (Density estimation) การเรียนรู้คุณลักษณะของข้อมูล (Feature learning) การลดมิติข้อมูล (Dimensionality reduction) และการใช้กฎความสัมพันธ์ของข้อมูล (Association rule) เป็นต้น [1]

#### 2.1.2 การประมวลผลทางภาษาธรรมชาติ

ถูกพัฒนาขึ้นเพื่อใช้สำหรับการวิเคราะห์ข้อความรีวิว ความรู้สึกต่าง ๆ ที่เกิดขึ้นจากผู้ใช้ แล้วทำการจำแนกความรู้สึกออกเป็นลักษณะต่าง ๆ เช่น ความรู้สึกเชิงบวก (Positive), ความรู้สึกเชิงลบ (Negative) และ ความรู้สึกเชิงปกติ (Neutral) [3] ซึ่งสามารถแบ่งระดับ (Level) การวิเคราะห์ความรู้สึกจากการประมวลผลทางภาษา ออกมาเป็น 3 แบบ ดังนี้

1. ระดับเอกสาร (Document level) เป็นการวิเคราะห์ความรู้สึกข้อความในระดับเอกสาร ซึ่งเป็นลักษณะการวิเคราะห์แบบภาพใหญ่
2. ระดับประโยค (Sentence level) เป็นการวิเคราะห์ความรู้สึกข้อความในระดับประโยค ซึ่งแต่ในแต่ละเอกสาร (document) จะประกอบไปด้วยประโยคต่าง ๆ มากมาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ระดับความคิดเห็น (Aspect level) จะเป็นการวิเคราะห์แง่มุมต่าง ๆ ที่มีอยู่ในประโยค

หากจำแนกออกเป็นประเภทต่าง ๆ ของการวิเคราะห์ความรู้สึกจากการประมวลผลทางภาษา สามารถจำแนกออกเป็นส่วนหลักที่นิยมในปัจจุบันได้ดังนี้ [4]

1. การวิเคราะห์ความรู้สึกจากการประมวลผลผลภาษาแบบละเอียด (Fine-grain sentiment analysis) เป็นการวิเคราะห์ข้อความในลักษณะเอกลักษณ์ (Entity) ของข้อความ ซึ่งใช้บ่งบอกถึงคุณลักษณะของข้อมูล (Feature) ต่าง ๆ ของประโยค ประกอบไปด้วยเอกลักษณ์ของข้อมูลที่เราคาดหวัง (Target Entity) มากมายในประโยค และยังมีขั้นตอนของการสกัดคำศัพท์แง่บวกความคิดเห็น (Aspect term extraction), การวิเคราะห์ความรู้สึกจากคำศัพท์แง่บวกความคิดเห็น (Aspect term polarity)

2. การวิเคราะห์ความรู้สึกจากการประมวลผลผลภาษาแบบไม่ละเอียด (Coarse-grain sentiment analysis) เป็นการวิเคราะห์ข้อความในลักษณะครอบคลุมเพียงแค่ว่าหนึ่งประโยค หรือทั้งเอกสารเท่านั้น ขั้นตอนของการทำวิเคราะห์ความรู้สึกจากการประมวลผลผลภาษามุ่งเน้นที่ระดับเอกสาร และ ระดับประโยคมุ่งเน้นการวิเคราะห์ความรู้สึก (Polarity) ของคำในประโยค

โดยส่วนใหญ่แล้วงานวิจัยจะเน้นที่การวิเคราะห์ความรู้สึกจากการประมวลผลผลภาษา โดยใช้การเรียนรู้แบบมีผู้สอน อาจทำให้เห็นคุณลักษณะของการวิเคราะห์ได้น้อย สามารถแสดงได้เพียงแค่ความรู้สึกเชิงบวก เชิงลบเพียงอย่างเดียว วิธีการเรียนรู้แบบไม่มีผู้สอน จึงอาจแก้ไขปัญหานี้ได้ ช่วยเติมเต็มช่องว่างของการวิเคราะห์ข้อมูล เห็นคุณลักษณะต่าง ๆ ของข้อความได้ลึกยิ่งขึ้น งานวิจัยส่วนใหญ่เน้นที่การวิเคราะห์ภาษาอังกฤษ [3] จึงอาจพบเห็นตัวอย่างการวิเคราะห์ภาษาไทยได้น้อย ทำให้รูปแบบการวิเคราะห์มีความแปลกใหม่ แสดงได้ถึงจุดเด่นของประโยคได้มากยิ่งขึ้น และสามารถพัฒนาให้เกิดประโยชน์ได้อย่างต่อเนื่อง

### 2.1.3 วิเคราะห์คุณลักษณะความคิดเห็นจากการประมวลผลทางภาษา

เป็นการนำข้อความเข้าสู่ขั้นตอนการวิเคราะห์ความรู้สึกจากการประมวลผลทางภาษา ซึ่งมีการเรียนรู้ด้วยภาษาธรรมชาติ และปัญญาประดิษฐ์ ซึ่งให้ผลลัพธ์การวิเคราะห์ทางภาษา แสดงออกมาสอดคล้องกับกลุ่มแง่บวกความคิดเห็นที่เกี่ยวข้องกับสิ่งที่ได้กำหนดไว้ ตัวอย่างประโยคที่ใช้วิเคราะห์ “คอมพิวเตอร์มีราคาแพงจัง” ดังนั้น แง่มุมความคิดเห็นของคอมพิวเตอร์ สามารถแสดงเป็นเซตคำตอบได้ดังนี้ { ฮาร์ดแวร์, ซอฟต์แวร์, ราคา } แต่ละคำแง่บวกความคิดเห็นใช้แทนคุณลักษณะ (attribute) ของคำที่เราสนใจ จากประโยคตัวอย่างเราจะได้คุณลักษณะของคำ คือ

คอมพิวเตอร์ [5] ความนิยมของการวิเคราะห์ความรู้สึกจากการประมวลผลภาษาแบบละเอียดเพิ่มขึ้น เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อย่างมากในเมื่อไม่นานมานี้ หลายงานวิจัยได้ทำการวิเคราะห์ข้อความในลักษณะตัวอย่างข้างต้น นอกจากนี้เราสามารถวิเคราะห์ความรู้สึกเพิ่มเติมในเชิงบวก เชิงลบจากแง่มุมความคิดเห็นที่เราได้ผลลัพธ์ งานวิจัย SemEval- 2014 [6] ซึ่งได้ถูกตีพิมพ์ออกมา โดยเน้นที่โดเมน (Domain) ข้อมูล 2 ชนิด คือ คอมพิวเตอร์ และร้านอาหาร รูปแบบของการวิเคราะห์ความรู้สึกของแง่มุมความคิดเห็น ถูกกำหนดขึ้นดังนี้ โดยรูปแบบการสกัดคำศัพท์แง่มุมความคิดเห็น (Aspect Term Extraction) และการวิเคราะห์ความรู้สึกจากคำศัพท์แง่มุมความคิดเห็น (Aspect Term Polarity) ถือเป็นวิธีการวิเคราะห์ความรู้สึกจากการประมวลผลผลภาษาแบบไม่ละเอียด

การสกัดคุณลักษณะจากความคิดเห็น (Aspect Term Extraction) ต้องมีการกำหนดประโยคที่ใช้ในการวิเคราะห์ จากนั้นจึงทำการค้นหาคำในประโยคที่สามารถแทนเอกลักษณ์ของประโยคได้ เพื่อนำไปวิเคราะห์หาคำศัพท์แง่มุมความคิดเห็น ซึ่งเกี่ยวข้องกับคำที่ใช้แทนเอกลักษณ์ของประโยค ยกตัวอย่างเช่น การสกัดคำนามของประโยคโดยใช้ Rule-base ร่วมกับกฎไวยากรณ์ โดยต้องเป็นคำที่สามารถแทนคุณลักษณะของประโยคที่เกิดขึ้นได้ เป็นต้น ดังเช่นตัวอย่างรูปที่ 2.1

Original text: ร้านตั้งอยู่ตึกพานิจิต  
Pos tag: (ร้าน, NOUN), (ตั้งอยู่, VERB), (ตึก, NOUN), (พานิจิต, NOUN)  
Aspect: (ร้าน, NOUN), (ตึก, NOUN)

#### รูปที่ 2.1 ตัวอย่างการสกัดคุณลักษณะจากความคิดเห็น

การวิเคราะห์ความรู้สึกจากคุณลักษณะ (Aspect Term Polarity) เป็นวิเคราะห์ความรู้สึกในเชิงบวก, เชิงลบ, เชิงปกติ จากคำคุณลักษณะที่เกิดขึ้นในประโยค ที่ได้มาด้วยขั้นตอนการจำแนกคำคุณลักษณะในประโยค ดังข้างต้น ซึ่งในบางครั้งอาจมีความขัดแย้งระหว่างความรู้สึกเชิงบวก เชิงลบที่มีปริมาณเท่ากัน ยกตัวอย่างเช่น การวิเคราะห์ความรู้สึกคำนามในประโยคด้วยวิเคราะห์คะแนนจากคำในคลังคำศัพท์ที่ประกอบด้วยคำแสดงความรู้สึก เป็นต้น สามารถแสดงได้ดังรูปที่ 2.2

Original text: อาหารอร่อยมาก  
Pos tag: (อาหาร, NOUN), (อร่อย, ADJ), (มาก, ADJ)  
Aspect: (อาหาร, NOUN), (ตึก, NOUN)  
Polarity: (อาหาร: Aspect), (อร่อย: Positive +1), (มาก: Positive +1)

#### รูปที่ 2.2 ตัวอย่างการสกัดความรู้สึกจากคุณลักษณะ

การจำแนกประเภทคุณลักษณะ (Aspect Category Detection) ต้องมีการกำหนดเซตของแง่มุมความคิดเห็นไว้ล่วงหน้า เช่น ความสะอาด = {สกปรก, เศษผม, เนื้อดิบ} เป็นต้น แล้วทำการค้นหาคำที่เกิดขึ้นในประโยคว่ามีคำใดบ้างที่เป็นสมาชิกของเซตที่กำหนดไว้ ประโยคนั้นจะถูกจำแนกเอกสารนี้ด้วยตัวแทนของเซตคำคุณลักษณะ ดังเช่นตัวอย่างในรูปที่ 2.3 อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Original text: มีเศษผมเยอะมาก

Aspect: (เศษผม, NOUN)

Aspect category: ความสะอาด

### รูปที่ 2.3 ตัวอย่างการจัดประเภทคุณลักษณะ

การวิเคราะห์ความรู้สึกจากประเภทคุณลักษณะ (Aspect Category Polarity) เป็นการวิเคราะห์ความรู้สึกในเชิงบวก, เชิงลบ, เชิงปกติ จากประเภทคุณลักษณะ ที่ได้มาด้วยขั้นตอนการจำแนกประเภทคุณลักษณะ สามารถแสดงตัวอย่างได้ดังรูปที่ 2.4

Original text: มีเศษผมเยอะมาก

Pos tag: (เศษผม, NOUN), (เยอะ, ADJ), (มาก, ADJ)

Aspect: (เศษผม, NOUN)

Aspect category: ความสะอาด

Polarity: (ความสะอาด: Aspect), (เยอะ: Negative -1)

### รูปที่ 2.4 ตัวอย่างการวิเคราะห์ความรู้สึกจากประเภทคุณลักษณะ

#### 2.1.4 คำพ้องความหมาย (Synonyms)

สำหรับภาษาอังกฤษจะพบคำที่มีความหมายเดียวกัน แต่มีการแปลงรูปคำศัพท์ต่าง ๆ มากมาย เช่น “eat” “ate” “eaten” ทั้งสามคำนี้ให้ความหมายเดียวกับคำว่า “eat” เป็นต้น แต่ในภาษาไทยนั้นมีการใช้คำศัพท์คำพ้องความหมายเสมือน เช่น “รับประทาน” “เสวย” และ “กิน” เป็นต้น ประโยคที่เกิดขึ้นในบทวิจารณ์เราจึงต้องมีการใช้คลังคำศัพท์คำพ้องความหมาย เพื่อทำการแบ่งกลุ่มคำที่เกิดขึ้นในความหมายลักษณะเดียวกัน

#### 2.1.5 แบบจำลองภาษา N-Gram

เป็นรูปแบบของการแสดงคุณลักษณะของข้อความในประโยคที่มีโอกาสเกิดขึ้นร่วมกันจำนวน N คำ จนเกิดขึ้นเป็นประโยค [16] มักใช้สำหรับการจำแนกคำในขั้นตอนการเตรียมข้อมูลวิเคราะห์ข้อความบทวิจารณ์ โดยมีการตัดคำในประโยคออกจากกัน และนำมาพิจารณาร่วมกันทีละจำนวน N คำ โดยที่  $N = 1, 2, 3, \dots$  ในงานวิเคราะห์มีการกำหนดค่า N ที่แตกต่างกันไปตามความเหมาะสม ถ้าหากจำนวน  $N = 1$  จะถูกเรียกว่า unigram เป็นการพิจารณาทีละคำในประโยค สามารถแสดงตัวอย่างประโยคได้ดังรูป 2.5 ด้านล่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Original text: มีเศษผมเยอะมาก

1-Gram: มี / เศษ / ผม / เยอะ / มาก

## รูปที่ 2.5 ตัวอย่างการจำแนกคุณลักษณะของข้อความในรูปแบบ 1-gram (unigram)

รูปแบบการแสดงข้อความ 1-Gram (unigram) เป็นการแสดงรูปแบบคำในประโยคที่เกิดขึ้นทีละหนึ่งคำในประโยค สามารถแสดงตัวอย่างได้เป็น (มี), (เศษ), (ผม), (เยอะ), (มาก) เป็นต้น

รูปแบบการแสดงข้อความ 2-Gram (bigram) เป็นการแสดงรูปแบบคำในประโยคที่เกิดขึ้นทีละสองคำในประโยค สามารถแสดงตัวอย่างได้เป็น (มี, เศษ), (เศษ, ผม), (ผม, เยอะ), (เยอะ, มาก) เป็นต้น โดยทั่วไปแล้วการวิเคราะห์คุณลักษณะนิยมการใช้ bigram ในการพิจารณา

รูปแบบการแสดงข้อความ 3-Gram (trigram) เป็นการแสดงรูปแบบคำในประโยคที่เกิดขึ้นทีละสามคำในประโยค สามารถแสดงตัวอย่างได้เป็น (มี, เศษ, ผม), (เศษ, ผม, เยอะ), (ผม, เยอะ, มาก) เป็นต้น

### 2.1.6 วิธีการเชิงความถี่ (Frequency-based Approach)

สำหรับขั้นตอนในการสกัดคุณลักษณะเชิงความถี่ มีทฤษฎีที่เกี่ยวข้องต่าง ๆ ตามลำดับ ดังนี้ Term Frequency (TF), Inverse Document Frequency (IDF) และ Term Frequency-Inverse Document Frequency (TF-IDF) โดยมีจุดประสงค์เพื่อให้โมเดลทำงานได้ดียิ่งขึ้น

Term Frequency-Inverse Document Frequency หรือเรียกว่า TF-IDF เป็นวิธีการทางสถิติสำหรับค้นหาคำที่มีความสำคัญในเอกสาร Term Frequency หรือเรียก TF เป็นการแสดงถึงคำที่มีความสำคัญ แทนค่าด้วย  $t$  โดยเกิดจากการคำนวณจำนวนคุณลักษณะแต่ละคำที่ปรากฏในเอกสารเทียบกับจำนวนคุณลักษณะทั้งหมดในเอกสาร Inverse Document Frequency หรือเรียกว่า IDF ใช้แสดงถึงความสำคัญของคำคุณลักษณะในเอกสารนั้น แทนค่าด้วย  $t$  โดยเกิดจากค่าผกผันของจำนวนเอกสารทั้งหมดเทียบกับจำนวนเอกสารที่ปรากฏคุณลักษณะเฉพาะแต่ละคำ สมการของ TF, IDF และ TF-IDF แสดงได้ดังต่อไปนี้ [9]

$$TF(t) = \frac{\text{จำนวนคำคุณลักษณะเฉพาะ } t \text{ ในเอกสาร}}{\text{จำนวนคุณลักษณะเฉพาะทั้งหมดในเอกสาร}}$$

$$IDF(t) = \log_e \frac{\text{จำนวนเอกสารทั้งหมด}}{\text{จำนวนเอกสารที่พบคำคุณลักษณะ } t}$$

$$IDF(t) = TF(t) \times IDF(t)$$

ในการคำนวณค่าน้ำหนัก TF-IDF จะใช้ผลคูณระหว่าง TF(t) และ IDF (t) โดยผลลัพธ์ค่าน้ำหนักที่เกิดขึ้นสามารถใช้แสดงข้อมูลในรูปแบบของ Vector Space Model (VSM) เพื่อใช้ในงานวิเคราะห์

บทวิจารณ์ด้วยโมเดลการเรียนรู้ของเครื่อง แต่ในงานวิจัยนี้ผู้วิจัยได้ใช้ประโยชน์จากคำคุณลักษณะเอกสารนี้เป็นเอกสารที่เลือกมาเพื่อใช้ในการทำเหมืองข้อมูลเพื่อหาข้อสรุปเกี่ยวกับเนื้อหาที่ปรากฏในเอกสารนี้ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และค่าน้ำหนักได้จากการคำนวณ TF-IDF สำหรับใช้ในการสกัดคุณลักษณะเชิงความถี่ด้วย อัลกอริทึม ตัวอย่างการแสดงผลลัพธ์ กำหนดให้เอกสารประกอบด้วยคำในประโยคจำนวน จำนวน 200 คำ และในเอกสารจำนวน 200 คำ ปรากฏคำว่า “อรรอย” จำนวน 10 ครั้ง สามารถแสดงค่า Term Frequency (TF) ได้เป็น  $10/250 = 0.04$  กำหนดให้มีจำนวนเอกสาร 50,000 เอกสาร และมีจำนวน 500 เอกสาร ปรากฏคำว่า “อรรอย” สามารถแสดงค่า Inverse Document Frequency (IDF) ได้เป็น  $50000/500 = 100$  จากตัวอย่างผลลัพธ์ที่เกิดขึ้นสามารถแสดงค่าน้ำหนัก TF-IDF(อรรอย) ได้จากผลคูณของ Term Frequency และ Inverse Document Frequency จะได้  $0.04 \times 100 = 4$

### 2.1.7 วิธีการใช้กฎฐาน (Rule-based Approach)

ในภาษาไทยประโยคมักเน้นการสื่อความหมายระหว่างผู้พูด และผู้ฟังเป็นหลัก บางครั้งจึงไม่จำเป็นต้องมีครบทั้งโครงสร้างประโยค ประธาน กริยา และกรรม ก็สามารถถือเป็นประโยคได้ ผู้พูดและผู้ฟังสามารถเข้าใจความหมายได้เป็นอย่างดี [17] เช่น ผู้พูดถามว่า “เมื่อเข้าคุณมาช้าหรือเปล่า” ผู้ฟังตอบว่า “ไม่” ก็สามารถให้ความหมายแทนประโยคที่มีครบทั้งไวยากรณ์ได้ “ไม่ได้มาช้าครับ” เป็นต้น

ทางภาษาศาสตร์สามารถแบ่งการวิเคราะห์ทางภาษาที่เกิดขึ้นในประโยคได้ 2 แนวทาง คือ 1.การวิเคราะห์ทางภาษาด้วยกฎไวยากรณ์ (Rule-based) เป็นการวิเคราะห์ภาษาด้วยกฎไวยากรณ์ที่เกิดขึ้นในประโยค ตัวอย่างกฎไวยากรณ์ นามวลี (Noun phrase) เช่น นาม + อกรรมย่อย = กระทะใหญ่, นาม + ลักษณะนาม = สัมสามใบ, นาม + อกรรมย่อย + ลักษณะนาม = คนผอมสองคน เป็นต้น สามารถสกัดคุณลักษณะนามได้จากกฎนามวลี 2.การวิเคราะห์โครงสร้างทางภาษาที่เกิดขึ้นจริง (Corpus-based) ในการใช้ภาษาโครงสร้างต่าง ๆ อาจไม่เป็นไปตามไวยากรณ์ ต้องมีการใช้งานความรู้ที่ได้จากคลังข้อมูลการใช้ภาษาที่เกิดขึ้นจริงในประโยค พร้อมทั้งมีการใช้งานข้อมูลทางสถิติเพิ่มเข้าไปในบทไวยากรณ์ จึงอาจต้องมีการวิเคราะห์ทางภาษานอกเหนือจากกฎไวยากรณ์ที่เกิดขึ้นในประโยค เช่น กริยา + นาม + อกรรมย่อย = ใช้กระทะใหญ่ เป็นต้น สามารถสกัดคุณลักษณะนามได้จากกฎภาษาที่เกิดขึ้นจริง

## 2.2 งานวิจัยที่เกี่ยวข้อง

ในการศึกษาวิจัยเกี่ยวกับการวิเคราะห์ความคิดเห็นจะมุ่งเน้นที่การวิเคราะห์ความรู้สึกจากการประมวลผลภาษาแบบไม่ละเอียดเป็นส่วนใหญ่ อาจทำให้การวิเคราะห์ความคิดเห็นได้ไม่ละเอียดมากพอ การวิเคราะห์ความคิดเห็นแบบการวิเคราะห์ความรู้สึกจากการประมวลผลภาษาแบบละเอียดจึงเข้ามาเติมเต็มการแก้ไขปัญหาส่วนนี้ ทำให้เราเข้าใจความรู้สึกของประโยคที่มากกว่าระดับเอกสาร และ ระดับประโยค ขณะที่การวิเคราะห์ความรู้สึกจากการประมวลผลภาษาแบบละเอียด จะพบได้

น้อยกว่า โดยทั้งหมดของงานวิจัยภาษาอังกฤษ, ภาษาจีน, ภาษาHindในการทดลอง ส่วนภาษาไทย เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์เพื่อการเรียนเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ในทางอื่นไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นั้นอาจพบได้น้อยกว่า โดยสามารถแบ่งกลุ่มงานวิจัยที่เกี่ยวข้องออกเป็น 3 กลุ่ม คือ กลุ่มงานวิจัยเกี่ยวกับปัจจัยในการวิเคราะห์ร้านอาหาร กลุ่มงานวิจัยการวิเคราะห์ข้อความแบบ Aspect Term Sentiment กลุ่มงานวิจัยการวิเคราะห์ข้อความแบบ Aspect Category Sentiment สามารถสรุปได้ดังตารางที่ 2.1

งานวิจัย	ชื่องานวิจัย	วัตถุประสงค์	เทคนิคที่ใช้
2.2.1 กลุ่มงานวิจัยเกี่ยวกับปัจจัยในการวิเคราะห์ร้านอาหาร			
1.	Restaurant Quality and Customer Satisfaction [10]	ศึกษาปัจจัย และผลกระทบต่าง ๆ ของความพึงพอใจลูกค้าในประเทศชาอูดีอาระเบีย และมีการทดสอบสมมุติฐานที่เกิดขึ้น	1. Correlation matrix 2. Regression analysis
2.2.2 กลุ่มงานวิจัยการวิเคราะห์ข้อความทางภาษาศาสตร์			
1.	Thai sentence parsing using genetic programming [17]	ศึกษาวิธีการวิเคราะห์การกระจายคำที่ใช้หลักไวยากรณ์ของภาษาไทยและพจนานุกรม	1. โครงสร้างไวยากรณ์ทางภาษาศาสตร์ 2. โครงสร้างไวยากรณ์ที่เกิดขึ้นจริง
2.2.3 กลุ่มงานวิจัยการวิเคราะห์ข้อความแบบ Aspect Term Sentiment			
1.	Aspect-based Sentiment Analysis using Dependency Parsing [3]	ค้นหาแง่มุมความคิดเห็นและวิเคราะห์ความรู้สึกจากรีวิวภาพยนตร์ในภาษา Hindi โดยใช้วิธี Dependency parsing	1. Hindi SentiWordNet (HSWN) 2. Hindi Dependency Parser (HDP) 3. Pointwise Mutual Information Retrieval (PMI-IR)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

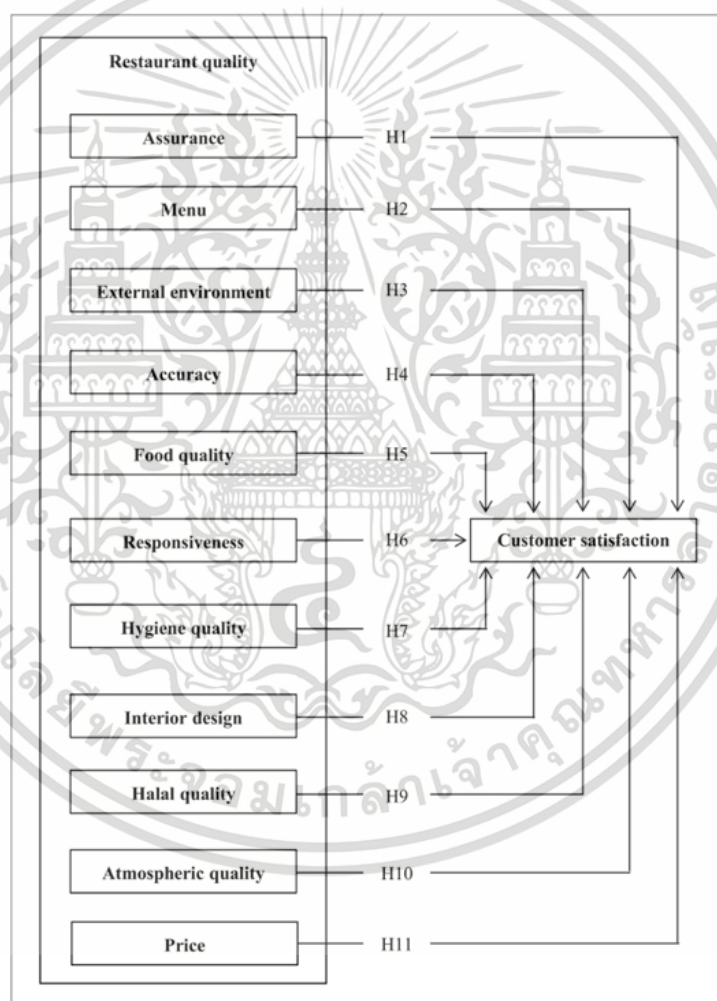
2.	Algorithm for Extracting Features from Product Reviews [15]	วิเคราะห์ความรู้สึกที่มีต่อคุณลักษณะบทวิจารณ์ผลิตภัณฑ์	1.Aspect Term Extraction (Frequency Approach & Rule-based Approach) 2.Aspect Term Sentiment 3.POS tagging
3.	Word Embedding-based Approach to Aspect Detection for Aspect-based Summarization of Persian Customer Reviews [18]	วิเคราะห์ข้อสรุปคุณลักษณะจากบทวิจารณ์ลูกค้าเปอร์เซียโดยประยุกต์ใช้ Word-Embedding	1.Word-Embedding 2.Calculation similarity vector 3.Aspect Categorization
2.2.4 กลุ่มงานวิจัยการวิเคราะห์ข้อความแบบ Aspect Category Sentiment			
1.	Aspect-Based Sentiment Analysis of Social Media Data With Pre-Trained Language Models [11]	วิเคราะห์แง่มุมความคิดเห็นของ social network โดยใช้ BERT แบบโมเดล Pre-training	1. RNN: LSTM & GRU 2. Bidirectional Encoder Representations with transformers (BERT)
2.	Unsupervised Aspect-Based Sentiment Analysis on Indonesian Restaurant Reviews[12]	วิเคราะห์แง่มุมความคิดเห็นของร้านอาหารในประเทศอินโดนีเซีย	1. Aspect Term Expansion algorithm
3.	A BERT-based Hierarchical Model for Vietnamese Aspect Based Sentiment Analysis [13]	วิเคราะห์แง่มุมความคิดเห็นของร้านอาหารและโรงแรมสำหรับภาษาเวียดนาม	1. Bidirectional Encoder Representations with transformers (BERT)

ตารางที่ 2.1 รายการงานวิจัยจำแนกตามหมวดหมู่  
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ภายในเพื่อการศึกษาเท่านั้น เมื่อผู้ใดเห็นไปประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.2.1 กลุ่มงานวิจัยเกี่ยวกับปัจจัยในการวิเคราะห์ร้านอาหาร

### 1. Restaurant Quality and Customer Satisfaction [10]

Bader M. A. Almohaimmeed (2014) ได้มีการศึกษาปัจจัย และผลกระทบต่าง ๆ ของความพึงพอใจลูกค้าที่มีต่อร้านอาหาร ในงานวิจัยนี้ได้ระบุการพิจารณา Restaurant quality ไว้ด้วยกัน 11 อย่าง ได้แก่ halal, food, hygiene, menu, atmospheric quality, assurance, accuracy, responsiveness, interior design, external environment และ price โดยพิจารณาร้านอาหารขนาดเล็ก ขนาดใหญ่ ของประเทศ Saudi Arabia จากงานวิจัยนี้ได้รวบรวมสมมุติฐานต่าง ๆ เกี่ยวกับร้านอาหารไว้ดังรูปที่ 2.6, 2.7 และ 2.8



รูปที่ 2.6 Restaurant Quality Hypothesis

(ที่มา: <https://www.econjournals.com/index.php/irmm/article/view/4452>)

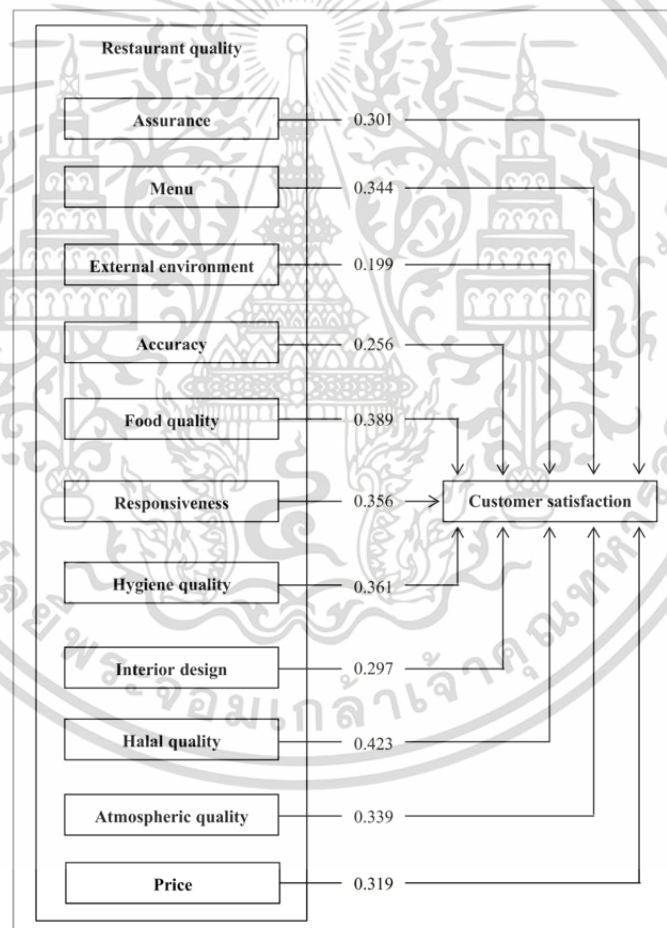
จากผลการทดลองด้วย regression analysis ในงานวิจัยชิ้นนี้ ให้ผลลัพธ์แสดงดังรูป 2.7 เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

H	Independent variable	Dependent variable	$\beta$	t	Result
H <sub>1</sub>	Assurance	Customer satisfaction	0.301*	4.201	Supported
H <sub>2</sub>	Menu		0.344*	6.115	Supported
H <sub>3</sub>	External environment		0.199*	4.001	Supported
H <sub>4</sub>	Accuracy		0.256*	4.120	Supported
H <sub>5</sub>	Food quality		0.389*	9.140	Supported
H <sub>6</sub>	Responsiveness		0.356*	4.212	Supported
H <sub>7</sub>	Hygiene quality		0.361*	6.320	Supported
H <sub>8</sub>	Interior design		0.297*	8.201	Supported
H <sub>9</sub>	Halal quality		0.423*	11.02	Supported
H <sub>10</sub>	Atmospheric quality		0.339*	4.332	Supported
H <sub>11</sub>	Price		0.319*	6.371	Supported

\*Significant at P<0.05

รูปที่ 2.7 Results of regression analysis (1)

(ที่มา: <https://www.econjournals.com/index.php/irmm/article/view/4452>)



รูปที่ 2.8 Results of regression analysis (2)

(ที่มา: <https://www.econjournals.com/index.php/irmm/article/view/4452>)

สามารถสรุปผลการทดลองของงานวิจัยชิ้นนี้ได้ว่า Halal quality มีผลต่อความพึงพอใจของลูกค้าเป็นอย่างมาก สำหรับการพิจารณาร้านอาหารในประเทศ Saudi Arabia อันดับถัดมา คือ Food quality เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.2.2 กลุ่มงานวิจัยการวิเคราะห์ข้อความทางภาษาศาสตร์

สำหรับงานวิจัยในกลุ่มนี้เป็นหลักการวิเคราะห์ข้อความทางภาษาศาสตร์โดยใช้วิธีการทางไวยากรณ์ และวิธีการทางสถิติ

### 1. Thai sentence parsing using genetic programming [17]

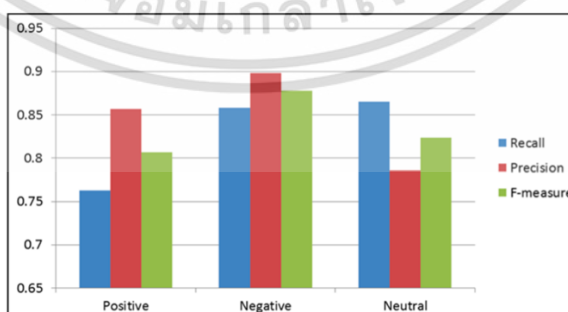
สมนึก สินธุปวน (2003) ได้มีการศึกษาการตัดคำ และการกระจายคำในภาษาไทย โดยพิจารณาถึงโครงสร้างไวยากรณ์ รวมทั้งมีการใช้วิธีการทางสถิติด้วยโปรแกรมเจนเนติกในการสร้างแผนภาพต้นไม้ตัดสัจใจเพื่อแสดงความเหมาะสมที่เกิดจากการตัดคำให้สอดคล้องกับไวยากรณ์มากที่สุด โดยรวมนั้นมีการใช้งานการวิเคราะห์ทางภาษาด้วยกฎไวยากรณ์ (Rule-based) และการวิเคราะห์ การวิเคราะห์โครงสร้างทางภาษาที่เกิดขึ้นจริง (Corpus-based)

## 2.2.3 กลุ่มงานวิจัยการวิเคราะห์ข้อความแบบ Aspect Term Sentiment

สำหรับงานวิจัยในกลุ่มนี้แง่มุมความคิดเห็นต่าง ๆ ที่เกิดขึ้น ต้องไม่มีการระบุไว้ล่วงหน้า ใช้วิธีการค้นหา knowledge จากข้อมูลที่มีอยู่ให้ได้ผลลัพธ์ออกมาสื่อความหมายมากที่สุด

### 1. Aspect-based Sentiment Analysis using Dependency Parsing [3]

SUJATA RANI และ PARTEEK KUMAR (2021) ได้มีการศึกษาการใช้งาน aspect-based sentiment analysis ร่วมกับภาษา Hindi เขาได้ใช้ Hindi Dependency Parser (HDP) เพื่อการระบุความสัมพันธ์ระหว่างแง่มุมความคิดเห็นที่เกิดขึ้นในประโยค และได้ใช้ Hindi SentiWordNet สำหรับการวิเคราะห์ข้อความ งานวิจัยนี้ได้มีการสร้าง dependency parser ระหว่างคำต่าง ๆ ในประโยค ทำให้สามารถระบุความสัมพันธ์ระหว่างคำที่เกี่ยวข้องกับแง่มุมความคิดเห็นที่ใกล้เคียงที่สุดได้ และคำนวณ polarity ที่เกี่ยวข้องกับแง่มุมความคิดเห็นโดยภาพรวมของประโยคได้ จากผลการทดลองให้ค่า accuracy 83.2%, precision 0.85, recall 0.83 และ F-measure 0.84 ซึ่งให้ผลดีกับดาต้าเซตรีวิวกาพยนตร์ แสดงได้ดังรูปที่ 2.9



รูปที่ 2.9 Accuracy of Hindi Dependency Parser

(ที่มา: <https://dl.acm.org/doi/10.1145/3485243>)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2. Algorithm for Extracting Features from Product Reviews [15]

ชนิตา แก้วเพชร (2020) จากงานวิจัยนี้ได้มีการพัฒนาอัลกอริทึมในการสกัดคุณลักษณะจากบทวิจารณ์ผลิตภัณฑ์ออนไลน์ภาษาอังกฤษ เพื่อให้ในการหาข้อสรุปของบทวิจารณ์ โดยใช้วิธีการ 3 แบบ ได้แก่ วิธีการสกัดคุณลักษณะผลิตภัณฑ์เชิงความถี่ วิธีการสกัดคุณลักษณะผลิตภัณฑ์ด้วยโครงสร้างไวยากรณ์ และวิธีการสกัดคุณลักษณะแบบไฮบริดที่มีการสกัดคุณลักษณะผลิตภัณฑ์เชิงความถี่ร่วมกับโครงสร้าง เป็นต้น พร้อมทั้งมีการประยุกต์ใช้งานกฎไวยากรณ์ที่ถูกสร้างขึ้นด้วยต้นไม้ตัดสินใจ และคลังคำศัพท์คำพ้องความหมาย สำหรับการจำแนกกลุ่มคุณลักษณะที่มีความหมายเหมือนกัน ข้อมูลที่นำมาทดลองได้ให้ผู้เชี่ยวชาญระบุคุณลักษณะที่เกี่ยวข้องในบทวิจารณ์แต่ละประเภท พร้อมทั้งระบุความรู้สึกที่ถูกตัดจากบทวิจารณ์นั้น เพื่อใช้สำหรับขั้นตอนการประเมินประสิทธิภาพ

## 3. Word Embedding-based Approach to Aspect Detection for Aspect-based Summarization of Persian Customer Reviews [18]

Seyyed Aref Razavi, Masoud Asadpour (2017) งานวิจัยนี้ได้มีการสกัดคุณลักษณะจากบทวิจารณ์ออนไลน์ทั้งหมด พร้อมทั้งมีการจัดหมวดหมู่คุณลักษณะที่มีความคล้ายคลึงกันเข้าด้วยกัน โดยคุณลักษณะนั้นอาจมีทั้งสื่อความหมายได้ และมีความหมายเป็นนัยใกล้เคียง สามารถสรุปวิธีการต่างๆจากงานวิจัยได้ดังนี้ 1.สร้างอัลกอริทึมในการสกัดคุณลักษณะ ด้วยการคิดคะแนนวลี (phrasal score) เพื่อใช้ในการพิจารณาคุณลักษณะที่เกิดคู่กัน โดยคะแนนวลีใช้สำหรับการคำนวณความสัมพันธ์ และความเกี่ยวเนื่องทางความหมายระหว่างคำ คุณลักษณะที่เกิดคู่กันคำนวณได้จาก Pointwise Mutual Information-Retrieval (PMI-IR) และค่าความคล้ายกันระหว่างเวกเตอร์ 2.สร้างกฎในการค้นหาคำที่เป็นคุณลักษณะที่เหมาะสม ทำให้สามารถจัดหาคำที่ไม่เกี่ยวข้องที่เกิดขึ้นจำนวนมากได้ เช่น ชื่อสินค้า แปรนตร์สินค้า เป็นต้น 3.จัดเตรียม word embedding สำหรับคำในบทวิจารณ์ เพื่อใช้สำหรับการขั้นตอนหาคำสำคัญในการเป็นคุณลักษณะ และการจัดกลุ่มคำคุณลักษณะ ซึ่งผู้วิจัยได้ชี้แจงว่าขั้นตอนเหล่านี้สามารถนำไปประยุกต์ใช้กับภาษาอื่น ๆ ได้อย่างง่ายดาย

### 2.2.4 กลุ่มงานวิจัยการวิเคราะห์ข้อความแบบ Aspect Category Sentiment

ในขอบเขตที่ได้กำหนดไว้ในตอนเริ่มต้น เช่น การกำหนดแง่มุมความคิดเห็นเกี่ยวกับร้านอาหาร เช่น Food quality, Service และ Environment โมเดลในงานวิจัยกลุ่มนี้จะทำการค้นหาเอกลักษณ์ต่าง ๆ ในข้อความ แล้วจึงแปลผลความสอดคล้องให้อยู่ในเซตของแง่มุมความคิดเห็นที่ได้กำหนดไว้ในช่วงเริ่มต้น วิธีการที่ใช้ในงานวิจัยมีทั้งการเรียนรู้แบบมีผู้สอน และการเรียนรู้แบบไม่มีผู้สอน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 1. Aspect-Based Sentiment Analysis of Social Media Data With Pre-Trained Language Models [11]

Anina Troya, Reshmi Gopalakrishna Pillai, Cristian Rodriguez Rivero, Zulkuf Genc, Subhradeep Kayal, Dogu Araci (2021) ได้มีการศึกษาการวิเคราะห์ข้อความจาก Social Network ด้วยวิธีการ Aspect-based ด้วยโมเดล BERT ซึ่งเป็นโมเดลภาษา transfer learning ที่มีการฝึกหัดไว้แล้ว งานวิจัยนี้เน้นที่การวิเคราะห์ข้อความหัวข้อ plant-based food โดยการใช้ Aspect-based sentiment analysis ข้อมูลที่ใช้วิเคราะห์นำมาจาก Twitter, SemEval 2014, Sentihood และการใช้งานไลบรารี Snsrape ผู้เขียนได้ใช้โมเดล LSTM และ GRU เพื่อเปรียบเทียบผลลัพธ์ที่ได้จากโมเดล BERT จากผลการทดลองพบว่า โมเดล BERT ที่มีการทำ Aspect Category Detection ให้ค่า accuracy 76.8 %, Aspect Sentiment Classification ให้ค่า accuracy 66 % สำหรับโมเดล LSTM, GRU มีการ error ระหว่างขั้นตอนการเทรน

### 2. Unsupervised Aspect-Based Sentiment Analysis on Indonesian Restaurant Reviews [12]

Dhanang Hadhi Sasmita, Alfani F. Wicaksono, Samuel Louvan และ Mirna Adriani (2017) ได้เสนองานวิจัยการวิเคราะห์รีวิวความรู้สึกที่มีต่อร้านอาหารในภาษาอินโดนีเซียแบบไม่มีผู้สอน สำหรับขั้นตอนการดำเนินงานมี 2 ส่วน 1) feature extraction 2) aspect sentiment orientation classification ในระยะเริ่มต้นได้มีการกำหนดข้อความระหว่าง aspect-sentiment ไว้เป็นคู่ล่วงหน้า เช่น เช็ตของแง่มุมความคิดเห็น {food, price, ambience, service} และ เช็ตของประเภทความรู้สึก {-1,+1} ระหว่างการทำงาน จะมีการหาค่า Top-N similar words สำหรับเทียบคำที่มีลักษณะความหมายคล้ายกัน

### 3. A BERT-based Hierarchical Model for Vietnamese Aspect Based Sentiment Analysis [13]

Oanh Thi Tran และ Viet The Bui (2020) ได้เสนองานวิจัยการวิเคราะห์ข้อความภาษาเวียดนามโดยใช้โมเดล BERT ดาต้าเซตนำมาจาก SemEval 2016 ที่มีทั้งโดเมนร้านอาหาร และ โรงแรม งานวิจัยนี้แบ่งการทำงานโมเดลออกเป็น 2 ส่วน ส่วนแรก การหาความสัมพันธ์ระหว่าง entity-aspect ส่วนที่สอง การหา polarity ของแง่มุมความคิดเห็นในการวัดประสิทธิภาพของโมเดลนั้น ทางผู้วิจัยได้เปรียบเทียบระหว่างโมเดล BERT กับโมเดลต่าง ๆ เช่น SVM, Multilayer Perceptron, Linear SVM และ CNN-base

Domain	Methods	Phase A (Entity&Aspect)			Phase B (Aspect polarity)		
		Pre	Rec	F1	Pre	Rec	F1
RESTAURANT	<i>Existing methods</i>						
	SA1: SVM-rich features	0.79	0.76	0.77	0.62	0.60	0.61
	SA2: MultiLayer Perceptron	0.88	0.38	0.54	0.79	0.35	0.48
	SA3: Linear SVM	0.62	0.62	0.62	0.52	0.52	0.52
	SA4: CNN-based	0.8475	0.7648	0.8040	–	–	–
	<i>Our methods</i>						
	mBERT HM	0.8403	0.8264	0.8333	0.6570	0.7143	0.6845
viBERT HM	84.21	84.25	<b>0.8423</b>	69.75	72.92	<b>0.7130</b>	

รูปที่ 2.10 Results of restaurant domain analysis

(ที่มา: <https://ieeexplore.ieee.org/abstract/document/9287650>)

Domain	Methods	Phase A (Entity&Aspect)			Phase B (Aspect Polarity)		
		Pre	Rec	F1	Pre	Rec	F1
HOTEL	<i>Existing methods</i>						
	SA1: SVM-rich features	0.76	0.66	0.70	0.66	0.57	0.61
	SA2: MultiLayer Perceptron	0.85	0.42	0.56	0.80	0.39	0.53
	SA3: Linear SVM	0.83	0.58	0.68	0.71	0.49	0.58
	SA4: CNN-based	0.8235	0.5975	0.6925	–	–	–
	<i>Our methods</i>						
	mBERT HM	0.8532	0.7604	0.8042	0.7817	0.6556	0.7131
viBERT HM	0.8393	0.8026	<b>0.8206</b>	0.8004	0.7001	<b>0.7469</b>	

รูปที่ 2.11 Results of hotel domain analysis

(ที่มา: <https://ieeexplore.ieee.org/abstract/document/9287650>)

จากงานวิจัยนี้พบว่ารูปที่ 2.10 และ 2.11 ทั้งดาด้าเซตจากโดเมนร้านอาหาร และโรงแรมนั้น เมื่อนำมาทดสอบกับโมเดล mBERT และ viBERT ให้ผลการทดลองที่ดีกว่าโมเดล SVM, Multilayer Perceptron, Linear SVM และ CNN-base

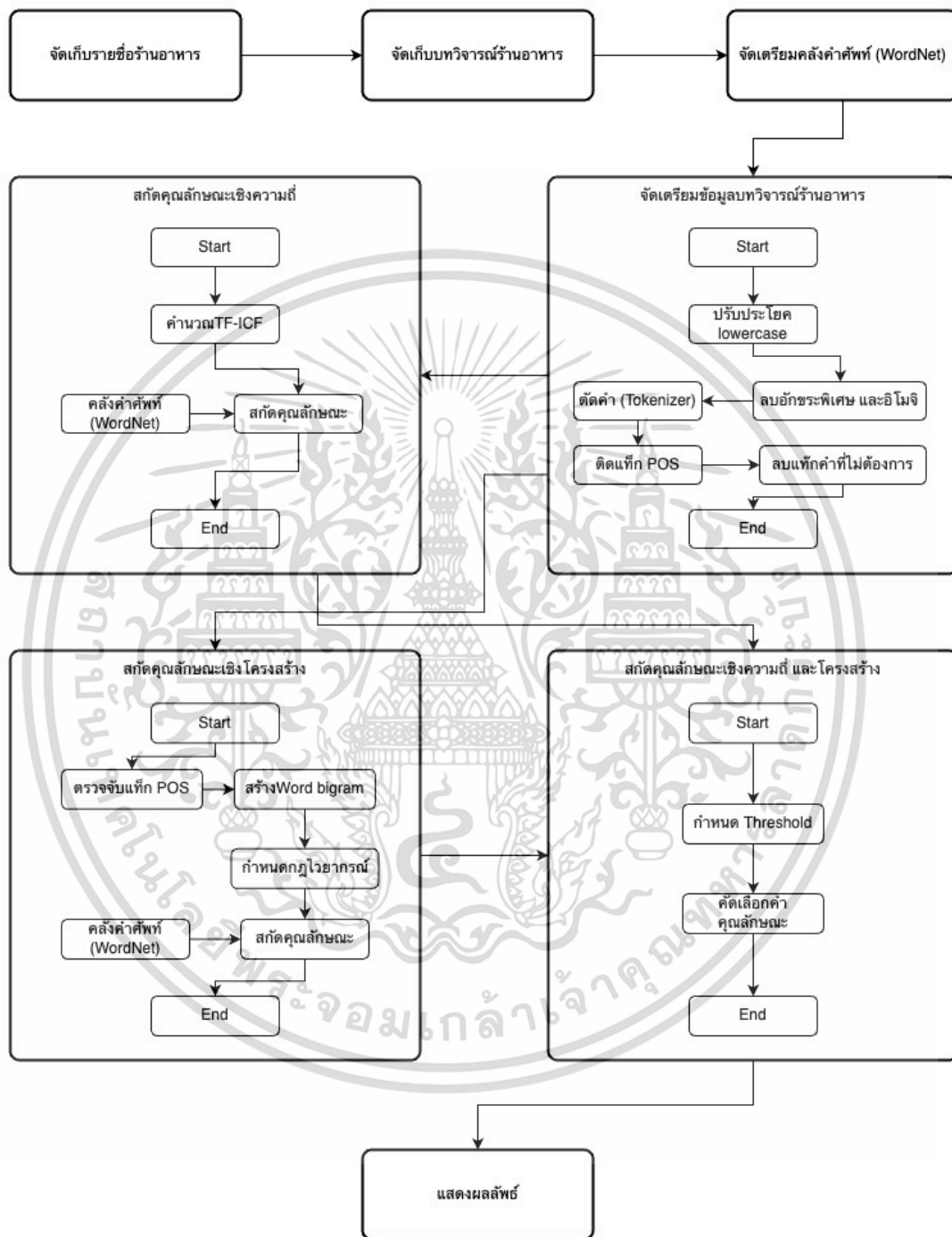
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### บทที่ 3

## วิธีดำเนินงานวิจัย

การดำเนินงานวิจัยเพื่อการสกัดคุณลักษณะความคิดเห็นร้านอาหารที่มีต่อการใช้บริการ เป็นกระบวนการสกัดคุณลักษณะที่บ่งบอกถึงความหมายสำคัญของประโยค ซึ่งถือเป็นกระบวนการสำคัญในการพิจารณาเลือกใช้บริการร้านอาหาร ช่วยให้ผู้ใช้บริการทราบถึงจุดเด่นของร้านที่ตรงใจแก่ผู้ใช้บริการ และนอกจากนี้ยังทำให้ผู้ประกอบการสามารถตอบสนองการให้บริการได้ตรงใจลูกค้าได้อย่างถูกต้อง ส่งผลให้ผู้ประกอบการได้รับความพึงพอใจจากลูกค้าเพิ่มขึ้น โดยผู้วิจัยได้นำเสนอการดำเนินงานวิจัยเพื่อการสกัดคุณลักษณะความคิดเห็น แสดงได้ดังรูปที่ 3.1 ซึ่งเป็นแผนภาพแสดงการทำงานโดยรวมทั้งหมด เริ่มจากการจัดเก็บรายชื่อร้านอาหารบนแพลตฟอร์มออนไลน์ การจัดเก็บรายชื่อร้านอาหารบนแพลตฟอร์มออนไลน์ การจัดเก็บบทวิจารณ์ร้านอาหารแต่ละร้าน การจัดเตรียมคลังข้อมูลคำพ้องความหมาย (WordNet) การจัดเตรียมข้อมูลบทวิจารณ์ร้านอาหารออนไลน์ การสกัดคุณลักษณะของผลิตภัณฑ์เชิงความถี่ การสกัดคุณลักษณะของผลิตภัณฑ์เชิงโครงสร้าง การสกัดคุณลักษณะของผลิตภัณฑ์เชิงความถี่ และโครงสร้าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.1 แผนภาพการทำงานของอัลกอริทึมสกัดคำคุณลักษณะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.1 การจัดเก็บรายชื่อร้านอาหารบนแพลตฟอร์มออนไลน์

ขั้นตอนการจัดเก็บบนวิจารณ์ออนไลน์ ผู้วิจัยได้ดำเนินเทคนิค web scraping เพื่อดึงข้อมูลจากหน้าเว็บไซต์ตามรูปแบบที่ได้กำหนด จากบทความร้านอาหารยอดนิยม ซึ่งผู้ใช้ได้ทำการจัดเก็บข้อมูลบทวิจารณ์จำนวน 8 ร้านค้า สามารถนำรายชื่อร้านอาหารเหล่านี้ใช้ในขั้นตอนการจัดเก็บบทวิจารณ์ร้านอาหารแต่ละร้าน โดยสามารถแจกแจงข้อมูลตัวอย่างได้ดังตารางที่ 3.1

Index	Title	Location	Link	Rating	Review_count
1	ร้านอาหาร 1	ทองหล่อ	https://www.food.com	5	768
2	ร้านอาหาร 2	อารีย์	https://www.food.com	4.1	1,924
3	ร้านอาหาร 3	สีลม	https://www.food.com	4.1	544
4	ร้านอาหาร 4	พระราม 9	https://www.food.com	4.2	768
5	ร้านอาหาร 5	สยามพารากอน	https://www.food.com	4	967
6	ร้านอาหาร 6		https://www.food.com	4.1	345
7	ร้านอาหาร 7	เสาชิงช้า	https://www.food.com	3	652

ตารางที่ 3.1 ตัวอย่างการจัดเก็บรายชื่อร้านอาหารบนแพลตฟอร์มออนไลน์

รายชื่อร้านอาหารที่จัดเก็บได้เบื้องต้นนั้น ข้อมูลบางชนิดไม่มีการระบุพิกัดข้อมูล จึงทำให้ข้อมูลบางส่วนจึงกลายเป็นค่าว่าง ซึ่งผู้วิจัยต้องจัดการกับค่าว่างเหล่านั้น

### 3.2 การจัดเก็บบทวิจารณ์ร้านอาหารแต่ละร้าน

การจัดเก็บบทวิจารณ์ร้านอาหารแต่ละร้านได้นำข้อมูลรายชื่อร้านอาหารจากขั้นตอน 3.1 ก่อนหน้านี้ มาทำการเก็บข้อมูลบทวิจารณ์เพิ่มเติมที่มีรายละเอียดมากขึ้น โดยใช้เทคนิค web scraping เช่นเดียวกัน ผู้วิจัยได้จัดเก็บในช่วง มกราคม 2022 บันทึกข้อมูลด้วยรูปแบบไฟล์ comma-separated values file (csv) ดังตาราง 3.2

Index	Date	Name	Title	Description	Rating
1	จ., 18 มิ.ย. 2018 12:36:29	EakSarawut	กลับไปทาน 3 วันติดต่อกัน	ร้านนี้มีที่จอดรถเยอะ มาก และรสชาติอาหาร อร่อย	5
2	ส., 8 ก.ย. 2018 12:34:29	WitTharit	เนื้อเยิ้มมาก	เนื้อนุ่มมาก คุ่มค่า สม ราคาดี น้ำจิ้มไม่เหมือน ใคร	4.1
3	ส., 2 มิ.ย. 2018 20:52:26	art	อร่อยดี ชอบ	พนักงานบริการดีมาก รสชาติอาหารก็ดีเช่นกัน	4.1

ตารางที่ 3.2 ตัวอย่างการจัดเก็บบทวิจารณ์ร้านอาหารแต่ละร้าน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.3 การจัดเตรียมคลังข้อมูลคำพ้องความหมาย (WordNet)

ผู้วิจัยได้นำคลังข้อมูลคำพ้องความหมายมาใช้ในการทำงาน [14] เพื่อใช้สำหรับการจำแนกคำคุณลักษณะที่ค้นพบในบทวิจารณ์ ในขั้นตอนของการสกัดคุณลักษณะเชิงความถี่ และคุณลักษณะเชิงโครงสร้าง เพื่อให้ได้ผลลัพธ์ที่เป็นคำคุณลักษณะที่มีความหมายเข้าใจง่ายมากยิ่งขึ้น โดยคลังข้อมูลนี้ประกอบไปด้วยคำภาษาไทยจำนวน 26242 คำ และถูกจำแนกกลุ่มความหมายจำนวน 498 กลุ่ม แสดงตัวอย่างได้ดังตารางที่ 3.3

ID	English	Thai	Gross
00613393 v	Abandon	ทิ้ง	stop maintaining or insisting on
00613393 v	Abandon	ละทิ้ง	stop maintaining or insisting on
00613393 v	Give up	เลิก	stop maintaining or insisting on
00613393 v	Give up	ยกเลิก	stop maintaining or insisting on
05200169 n	Ability	ความสามารถ	the quality of being able to perform
05200169 n	Ability	พรสวรรค์	the quality of being able to perform

ตารางที่ 3.3 ตัวอย่างการจัดเก็บคลังข้อมูลคำพ้องความหมาย

### 3.4 การจัดเตรียมข้อมูลบทวิจารณ์ร้านอาหารออนไลน์

การจัดเตรียมบทวิจารณ์ร้านอาหารแต่ละร้านผู้วิจัยได้นำข้อมูลจากแต่ละร้านอาหารมาทำการปรับปรุงคุณภาพของข้อมูล เพื่อให้ได้ผลลัพธ์สำหรับการนำไปสกัดคุณลักษณะของข้อมูลดียิ่งขึ้น ขั้นตอนการทำงานประกอบไปด้วย 5 ขั้นตอน ดังนี้ การปรับข้อความเป็นตัวพิมพ์เล็ก (lower case) การลบอักขระพิเศษ และอีโมจิ การกำหนดค่าโมเดลการตัดคำ (Tokenizer) การติดแท็ก Part-Of-Speech (POS) ในประโยค การขจัดแท็ก POS ที่ไม่จำเป็นในประโยค

**3.4.1 การปรับข้อความเป็นตัวพิมพ์เล็ก (lower case)** ขั้นตอนนี้ผู้วิจัยได้ทำการปรับตัวอักษรบางชนิดที่เป็นตัวพิมพ์ใหญ่ให้กลายเป็นตัวพิมพ์เล็ก เพื่อให้ข้อมูลมีความสะอาดมากยิ่งขึ้น ง่ายต่อการใช้งานโมเดลการตัดคำ (Tokenizer) สามารถแสดงได้ดังรูปที่ 3.2

Original text: เมนูที่แนะนำ Special menu Tokujyo Mitsuji  
 Preparing text: เมนูที่แนะนำ special menu tokujyo mitsuji

รูปที่ 3.2 ตัวอย่างการปรับข้อความเป็นตัวพิมพ์เล็ก

**3.4.2 การลบอักขระพิเศษ และอีโมจิ** ข้อความบทวิจารณ์ร้านอาหารประกอบไปด้วยอักขรหึ่งไทย และอังกฤษ ซึ่งมีทั้งตัวอักขระพิเศษ และอีโมจิ เช่น "#\()\*,-;<=> 😊 😄 😁 เป็นต้น จึงต้องทำการขจัดออก เพื่อให้อัลกอริทึมการสกัดคุณลักษณะเชิงความถี่ และเชิงโครงสร้าง ทำงานได้ดียิ่งขึ้น รูปที่ 3.3 แสดงตัวอย่างการขจัดอักขระพิเศษ และอีโมจิ

Original text: ข้าวผัดกระเทียม (แพงมาก) วันที่ไปข้าวญี่ปุ่นหมด 😊  
 Preparing text: ข้าวผัดกระเทียมแพงมาก วันที่ไปข้าวญี่ปุ่นหมด

รูปที่ 3.3 ตัวอย่างการขจัดอักขระพิเศษ และอีโมจิ

**3.4.3 การกำหนดค่าโมเดลการตัดคำ (Tokenizer)** โมเดลการตัดคำนั้นจะเป็นการนำประโยคของบทวิจารณ์ มาทำการตัดออกแต่ละส่วน ให้กลายเป็นคำย่อยในประโยค เพื่อให้ได้ผลลัพธ์สำหรับการนำไปใช้ขั้นตอนการสกัดคุณลักษณะเชิงความถี่ ผู้วิจัยได้ใช้โมเดลการตัดคำจาก pythainlp โดยขั้นตอนของการตัดคำนั้น ได้มีการเพิ่มคำที่พบได้ทั่วไปในประโยคภาษาไทย (Thai stop word) เช่น ไว้ ไม่ ไป ได้ ให้ ไน โดย เป็นต้น เพื่อให้การตัดคำมีการขจัดคำเหล่านี้ออกไป นอกจากนี้ยังได้มีการเพิ่มคำเฉพาะ (Custom word) ที่พบบ่อยในบทวิจารณ์ร้านอาหาร เช่น ทองหล่อ, ร้านอาหาร, เนื้อย่าง, ปิ้งย่าง, หม้อหิ้น เป็นต้น เพื่อไม่ให้มีการตัดคำเฉพาะเหล่านี้ออกจากกัน ในประโยค ผลลัพธ์จากการตัดคำ แสดงดังรูปที่ 3.4

Original text: กินเนื้อย่างได้  
 Preparing text (Tokenizer): กิน เนื้อ ย่าง ได้  
 Preparing text (Tokenizer + Stop word (ได้)): กิน เนื้อ ย่าง  
 Preparing text (Tokenizer + Stop word(ได้) + Custom word(เนื้อย่าง)): กิน เนื้อย่าง

รูปที่ 3.4 ตัวอย่างการใช้งานโมเดลการตัดคำ

**3.4.4 การติดแท็ก Part-Of-Speech (POS)** เป็นการติดแท็กชนิดคำจากโครงสร้างภาษา และบริบทของประโยค ผู้วิจัยได้ใช้ไลบรารีจาก Pythainlp สำหรับการดำเนินงานขั้นตอนนี้ และได้เลือกใช้คลังคำศัพท์ (corpus) ของ Thai Parallel Universal Dependencies (Thai-PUD) ซึ่งได้มีแท็กต่างๆเกิดขึ้นมากมายประกอบด้วย ADJ, ADP, ADV, AUX, CCONJ, DET, NOUN, NUM, PART, PRON, PROPN, PUNCT, SCONJ, SYM, VERB เป็นต้น แสดงดังรูปที่ 3.5

Original text: ร้านตั้งอยู่ตึกพานิจิต  
 Pos tag: (ร้าน, NOUN), (ตั้งอยู่, VERB), (ตึก, NOUN), (พานิจิต, NOUN)

รูปที่ 3.5 ตัวอย่างการแท็ก POS

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**3.4.5 การขจัดแท็กชนิดคำที่ไม่จำเป็น** ขั้นตอนนี้ผู้วิจัยได้ทำการลบแท็กของคำบางชนิดที่ไม่เกี่ยวข้องกับการวิเคราะห์ เช่น SYM : Symbol สัญลักษณ์ต่างๆ, PUNC : Punction เครื่องหมายวรรคตอน และ NUM : Number ตัวเลข ช่วยให้คำที่ซ้ำซ้อนมีระเบียบมากยิ่งขึ้น

### 3.5 การสกัดคุณลักษณะของผลิตภัณฑ์เชิงความถี่

ขั้นตอนการสกัดคุณลักษณะของผลิตภัณฑ์เชิงความถี่ประกอบด้วยขั้นตอน 2 วิธี ได้แก่ การหาค่า Term frequency - Inverse Document Frequency (TF-IDF) การค้นหาคำแทนคุณลักษณะ เป็นต้น [15]

**3.5.1 การหาค่า Term frequency - Inverse Document Frequency (TF-IDF)** จากอัลกอริทึมดังกล่าวข้างต้นนี้จะใช้การคำนวณองค์ประกอบของคำแต่ละคำเทียบกับคำในเอกสารทั้งหมด แต่ทางผู้วิจัยได้มีประยุกต์ใช้การคำนวณองค์ประกอบของคำแต่ละคำโดยคิดจากคำในบทวิจารณ์ร้านอาหารทั้งหมด เรียกว่า Term frequency - Inverse Comment Frequency (TF - ICF) การคำนวณนั้นถูกแบ่งออกเป็น 2 ส่วน ได้แก่ การคำนวณ Term Frequency (TF), การคำนวณ Inverse Comment Frequency (ICF) และการคำนวณผลคูณของ Term Frequency - Inverse Comment Frequency

การคำนวณ Term Frequency คือ จำนวนของคำที่สามารถพบได้ในความคิดเห็น ซึ่งคำที่พบในความคิดเห็นได้มากที่สุด อาจสื่อถึงใจความสำคัญของบทวิจารณ์มาก ยกตัวอย่างเช่น หากเราพิจารณาบทวิจารณ์ร้านอาหาร เราอาจพบเจอคำว่า ‘อาหาร’, ‘รสชาติ’ เป็นจำนวนมาก โดยการคำนวณ Term Frequency คิดได้จากจำนวนของคำคุณลักษณะแต่ละคำเทียบกับจำนวนคำทั้งหมดในบทวิจารณ์ แสดงได้ดังสมการด้านล่างนี้

$$TF(t, c) = \frac{\text{number of terms } t \text{ appears in } c}{\text{total number of comments in } c}$$

กำหนดให้ **t** แทนค่า คำคุณลักษณะในบทวิจารณ์,

กำหนดให้ **c** แทนค่า บทวิจารณ์ร้านอาหาร

ตัวอย่างการคำนวณสามารถแสดงตารางที่ 3.4 และตารางที่ 3.5

บทวิจารณ์ที่ 1: ร้านนี้ทำอาหารได้อร่อยมาก รสชาติกลมกล่อมถูกปากทุกเมนู อาหารสะอาด

คำ	จำนวนคำ	TF value
ร้าน	1	1/14
นี้	1	1/14
ทำ	1	1/14
อาหาร	2	2/14

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรณีใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านอื่นๆ

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำ	จำนวนคำ	TF value
ได้	1	1/14
อ่อย	1	1/14
มาก	1	1/14
รสชาติ	1	1/14
ถูก	1	1/14
ปาก	1	1/14
ทุก	1	1/14
เมนู	1	1/14
สะอาด	1	1/14

ตารางที่ 3.4 การคำนวณค่า TF value บทวิจารณ์ที่ 1

บทวิจารณ์ที่ 2: ขึ้นชื่อเรื่องเนื้อ ปรั่งอ่อยมาก รสชาติเนื้อกลมกล่อม

คำ	จำนวนคำ	TF value
ขึ้น	1	1/11
ชื่อ	1	1/11
เรื่อง	1	1/11
เนื้อ	2	2/11
ปรั่ง	1	1/11
อ่อย	1	1/11
มาก	1	1/11
รสชาติ	1	1/11
กลม	1	1/11
กลม	1	1/11

ตารางที่ 3.5 การคำนวณค่า TF value บทวิจารณ์ที่ 2

การคำนวณ Inverse Comment Frequency คือ จำนวนของคำที่สามารถพบได้ในความคิดเห็น ซึ่งคำที่พบในความคิดเห็นได้มากที่สุด อาจสื่อถึงใจความสำคัญของบทวิจารณ์ได้น้อย ซึ่งตรงข้ามกับนิยามของ Term Frequency โดยการคำนวณ Inverse Comment Frequency คิดได้จากจำนวนบทวิจารณ์ทั้งหมด เทียบกับจำนวนของคำคุณลักษณะที่เกิดขึ้นในบทวิจารณ์ แสดงได้ดังสมการด้านล่างนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$ICF(t, c) = \log \left( \frac{\text{total number of comments in } c}{\text{number of comments appears in terms } t} \right)$$

กำหนดให้  $t$  แทนค่า คำคุณลักษณะในบทวิจารณ์,

กำหนดให้  $c$  แทนค่า บทวิจารณ์ร้านอาหาร

ตัวอย่างการคำนวณสามารถแสดงตารางที่ 3.6 และตารางที่ 3.7

บทวิจารณ์ที่ 1: ร้านนี้ทำอาหารได้อร่อยมาก รสชาติกลมกล่อมถูกปากทุกเมนู อาหารสะอาด

คำ	จำนวนบทวิจารณ์ที่ปรากฏ	IDF value
ร้าน	1	$\log\left(\frac{2}{1}\right) = 0.3010$
นี้	1	$\log\left(\frac{2}{1}\right) = 0.3010$
ทำ	1	$\log\left(\frac{2}{1}\right) = 0.3010$
อาหาร	1	$\log\left(\frac{2}{1}\right) = 0.3010$
ได้	1	$\log\left(\frac{2}{1}\right) = 0.3010$
อร่อย	2	$\log\left(\frac{2}{2}\right) = 0$
มาก	2	$\log\left(\frac{2}{2}\right) = 0$
รสชาติ	2	$\log\left(\frac{2}{2}\right) = 0$
ถูก	1	$\log\left(\frac{2}{1}\right) = 0.3010$
ปาก	1	$\log\left(\frac{2}{1}\right) = 0.3010$
ทุก	1	$\log\left(\frac{2}{1}\right) = 0.3010$
เมนู	1	$\log\left(\frac{2}{1}\right) = 0.3010$
สะอาด	1	$\log\left(\frac{2}{1}\right) = 0.3010$

ตารางที่ 3.6 การคำนวณค่า ICF value บทวิจารณ์ที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทวิจารณ์ที่ 2: ขึ้นชื่อเรื่องเนื้อ ปรุงอร่อยมาก รสชาติเนื้อกลมกล่อม

คำ	จำนวนบทวิจารณ์ที่ปรากฏ	IDF value
ขึ้น	1	$\log\left(\frac{2}{1}\right) = 0.3010$
ชื่อ	1	$\log\left(\frac{2}{1}\right) = 0.3010$
เรื่อง	1	$\log\left(\frac{2}{1}\right) = 0.3010$
เนื้อ	1	$\log\left(\frac{2}{1}\right) = 0.3010$
ปรุง	1	$\log\left(\frac{2}{1}\right) = 0.3010$
อร่อย	2	$\log\left(\frac{2}{2}\right) = 0$
มาก	2	$\log\left(\frac{2}{2}\right) = 0$
รสชาติ	2	$\log\left(\frac{2}{2}\right) = 0$
กลม	1	$\log\left(\frac{2}{1}\right) = 0.3010$
กลม	1	$\log\left(\frac{2}{1}\right) = 0.3010$

ตารางที่ 3.7 การคำนวณค่า ICF value บทวิจารณ์ที่ 2

การคำนวณ TF-ICF คือ ค่าที่แสดงถึงผลคูณของ Term Frequency และ Inverse Comment Frequency โดยค่าที่มากจะแสดงถึงคำที่มีความสำคัญสูงในแต่ละบทวิจารณ์ ซึ่งมีการกล่าวถึงบ่อย และไม่ปรากฏในบทวิจารณ์มากจนเกินไป สามารถแสดงสมการได้ดังนี้

$$TF - ICF = TF \times ICF$$

กำหนดให้ **TF** แทนค่า Term Frequency,

กำหนดให้ **ICF** แทนค่า Inverse Comment Frequency

ตัวอย่างการคำนวณสามารถแสดงตารางที่ 3.8 และตารางที่ 3.9

บทวิจารณ์ที่ 1: ร้านนี้ทำอาหารได้อร่อยมาก รสชาติกลมกล่อมถูกปากทุกเมนู อาหารสะอาด

คำ	TF value	ICF value	TF-ICF value
ร้าน	1/14	$\log\left(\frac{2}{1}\right) = 0.3010$	0.0215
นี้	1/14	$\log\left(\frac{2}{1}\right) = 0.3010$	0.0215
ทำ	1/14	$\log\left(\frac{2}{1}\right) = 0.3010$	0.0215

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้ใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำ	TF value	ICF value	TF-ICF value
อาหาร	2/14	$\log\left(\frac{2}{1}\right) = 0.3010$	<b>0.043</b>
ได้	1/14	$\log\left(\frac{2}{1}\right) = 0.3010$	0.0215
อร่อย	1/14	$\log\left(\frac{2}{2}\right) = 0$	0
มาก	1/14	$\log\left(\frac{2}{2}\right) = 0$	0
รสชาติ	1/14	$\log\left(\frac{2}{2}\right) = 0$	0
ถูก	1/14	$\log\left(\frac{2}{1}\right) = 0.3010$	0.0215
ปาก	1/14	$\log\left(\frac{2}{1}\right) = 0.3010$	0.0215
ทุก	1/14	$\log\left(\frac{2}{1}\right) = 0.3010$	0.0215
เมนู	1/14	$\log\left(\frac{2}{1}\right) = 0.3010$	0.0215
สะอาด	1/14	$\log\left(\frac{2}{1}\right) = 0.3010$	0.0215

ตารางที่ 3.8 การคำนวณค่า TF-ICF value บทวิจารณ์ที่ 1

บทวิจารณ์ที่ 2: ขึ้นชื่อเรื่องเนื้อ ปรั่งอร่อยมาก รสชาติเนื้อกลมกล่อม

คำ	TF value	ICF value	TF-ICF value
ขึ้น	1/11	$\log\left(\frac{2}{1}\right) = 0.3010$	0.0274
ชื่อ	1/11	$\log\left(\frac{2}{1}\right) = 0.3010$	0.0274
เรื่อง	1/11	$\log\left(\frac{2}{1}\right) = 0.3010$	0.0274
เนื้อ	2/11	$\log\left(\frac{2}{1}\right) = 0.3010$	<b>0.0547</b>
ปรั่ง	1/11	$\log\left(\frac{2}{1}\right) = 0.3010$	0.0274
อร่อย	1/11	$\log\left(\frac{2}{2}\right) = 0$	0
มาก	1/11	$\log\left(\frac{2}{2}\right) = 0$	0
รสชาติ	1/11	$\log\left(\frac{2}{2}\right) = 0$	0
กลม	1/11	$\log\left(\frac{2}{1}\right) = 0.3010$	0.0274
กลม	1/11	$\log\left(\frac{2}{1}\right) = 0.3010$	0.0274

ตารางที่ 3.9 การคำนวณค่า TF-ICF value บทวิจารณ์ที่ 2

จากตารางคำนวณ TF-ICF ข้างต้น ผู้วิจัยสามารถจำแนกใจความสำคัญของบทวิจารณ์ที่ 1 ออกมาได้เป็น “อาหาร” ส่วนบทวิจารณ์ที่ 2 สามารถจำแนกใจความสำคัญออกมาได้เป็น “เนื้อ” เอกสารนี้เป็นเอกสารที่สวอนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5.2 การค้นหาคำแทนคุณลักษณะโดยใช้เซตคำพ้องความหมาย สำหรับขั้นตอนนี้ผู้วิจัยได้ใช้คลังคำศัพท์ภาษาไทย [14] หรือที่เรียกว่า Wordnet ในการค้นหาคำคุณลักษณะ ร่วมกับการคำนวณค่า TF-ICF โดยคลังคำศัพท์คำพ้องความหมาย (Corpus:  $C_{sa}$ ) ประกอบไปด้วยคำพ้องความหมาย ( $W_{sa}$ ) สามารถแสดงได้เป็น  $W_{sa}(i) \in C_{sa}$  โดยที่  $i = 1, 2, 3, 4, \dots$  ซึ่งจะมีการค้นหาคำคุณลักษณะภายในประโยค ด้วยการใช้คำศัพท์ที่ได้มาจากการคำนวณ TF-ICF เพื่อนำมาจำแนกตามเซตคำพ้องความหมายแต่ละคำ สามารถแสดงได้เป็น  $W_i \in W_{sa}$  โดยที่  $i = 1, 2, 3, 4, \dots$  แต่ละเซตคำพ้องความหมาย คำคุณลักษณะที่มีค่า TF-ICF มากที่สุด จะถูกใช้เป็นตัวแทนประจำเซตคำพ้องความหมาย แผนภาพแสดงผลลัพธ์การจำแนกคำคุณลักษณะ สามารถแสดงได้ดังแผนภาพ Algorithm 1 และตารางที่ 3.10

---

**Algorithm 1** Extract Keywords From Wordnet

---

Input:  $C_{sa}$  (WordNet corpus),  $W_i$  (TF-ICF word)

Output:  $W_{sa}(i)$  (Corpus keyword)

---

```

1  for  $W_{sa}(i)$  in  $C_{sa}$ :
2    corpusList = []
3    for  $W_j$  in  $W_{sa}$ :
4      if  $W_j$  in  $W_{sa}(i)$ :
5        corpusList add  $W_j$ 
6    Term of  $W_{sa}(i)$  = max (TF-ICF in corpusList)
7    TF-ICF of  $W_{sa}(i)$  = mean (TF-ICF in corpusList)
8    Print("Next corpus")

```

$C_{sa}$	TF-ICF ของแต่ละคำ			คำที่ถูกเลือก
$W_{sa1}$	$W_1(0.35)$	$W_2(0.2)$	$W_3(0.5)$	$W_3$
$W_{sa2}$	$W_1(0.6)$			$W_1$
$W_{sa3}$	$W_1(0.35)$	$W_2(0.7)$		$W_2$
$W_{sa4}$	$W_1(0.4)$	$W_2(0.2)$	$W_3(0.02)$	$W_1$

ตารางที่ 3.10 การคัดเลือกคำแทนคุณลักษณะจาก TF-ICF ของแต่ละคำ

หลังจากผ่านขั้นตอนการคัดเลือกคำแทนคุณลักษณะแล้ว จะต้องมีการคำนวณค่า TF-ICF จากคำคุณลักษณะที่ได้มา และใช้ค่าเฉลี่ย TF-ICF แทนค่าคุณลักษณะเดิม ซึ่งคำนวณได้จาก ค่า TF-เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ICF ของคำคุณลักษณะภายในเซตคำพ้องความหมาย และจำนวนสมาชิกภายในเซตคำพ้องความหมาย ซึ่งสามารถแสดงการคำนวณได้ดังตัวอย่างดังนี้ กำหนดให้  $C_{sa} = \{W_{sa1}, W_{sa2}\} = \{Restaurant, Agree\}$

$$\begin{aligned} \text{เซตคำพ้องที่ 1: Restaurant} &= [W_1, W_2] \\ &= [\text{ร้านอาหาร}(0.35), \text{ห้องอาหาร}(0.26)] \\ &= [\text{ร้านอาหาร}(\frac{0.35+0.26}{2})] \\ &= [\text{ร้านอาหาร}(0.305)] \end{aligned}$$

ดังนั้น จากเซตคำพ้องความหมายที่ 1 “ร้านอาหาร” ( $W_1$ ) จึงถูกคัดเลือกให้เป็นตัวแทนของคำแทนคุณลักษณะ Restaurant และมีค่า TF-ICF ที่ได้จากการคำนวณค่าเฉลี่ย เท่ากับ 0.305

$$\begin{aligned} \text{เซตคำพ้องที่ 2: Agree} &= [W_1, W_2] \\ &= [\text{เข้ากัน}(0.05), \text{เหมาะสม}(0.57)] \\ &= [\text{เหมาะสม}(\frac{0.05+0.57}{2})] \\ &= [\text{เหมาะสม}(0.31)] \end{aligned}$$

ดังนั้น จากเซตคำพ้องความหมายที่ 2 “เหมาะสม” ( $W_2$ ) จึงถูกคัดเลือกให้เป็นตัวแทนคำแทนคุณลักษณะ Agree และมีค่า TF-ICF ที่ได้จากการคำนวณค่าเฉลี่ย เท่ากับ 0.31

### 3.6 การสกัดคุณลักษณะของผลิตภัณฑ์เชิงโครงสร้าง

การสกัดคุณลักษณะเชิงโครงสร้างนั้น ผู้วิจัยได้หยิบยกคำในประโยคต่างๆ ที่ได้ทำการติดแท็ก Part-Of-Speech จากขั้นตอนการเตรียมข้อมูลในตอนแรก ซึ่งได้มีแท็กต่างๆเกิดขึ้นมากมาย ประกอบด้วย ADJ, ADP, ADV, AUX, CONJ, DET, NOUN, NUM, PART, PRON, PROPN, PUNCT, SCONJ, SYM, VERB เป็นต้น ซึ่งหลังจากผ่านขั้นตอนนี้ ผู้วิจัยได้มีการพิจารณาคำในประโยค โดยพิจารณาจาก 2 คำรอบข้างที่ติดกัน (Bigram) [15] ด้วยจำนวนข้อมูลบทวิจารณ์ร้านอาหารที่มีมาก ทำให้เราสามารถแสดงตัวอย่างการจำแนกกฎไวยากรณ์ออกมาเบื้องต้นได้ ดังตารางที่ 3.11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กฎไวยากรณ์	ตัวอย่างคำ	คำคุณลักษณะ
NOUN+NOUN	(ร้าน, NOUN), (ปิ้งย่าง, NOUN)	ปิ้งย่าง
NOUN+VERB	(คนญี่ปุ่น, NOUN), (เลี้ยง, VERB)	คนญี่ปุ่น
VERB+NOUN	(คุ้ม, VERB) (ราคา, NOUN)	ราคา
NOUN+ADV	(บริการ, NOUN), (เร็ว, ADV)	บริการ
NOUN+ADJ	(เนื้อ, NOUN), (ธรรมดา, ADJ)	เนื้อ
NOUN+ADP	(บริการ, NOUN), (ทั่วถึง, ADP)	บริการ
NOUN+DET	(ราคา, NOUN), (นี้, DET)	ราคา
AUX+NOUN	(เป็น, AUX), (ญี่ปุ่น, NOUN)	ญี่ปุ่น
AUX+VERB	(ต้อง, AUX), (จอง, VERB)	จอง
ADV+ADP	(อ้อย, ADV), (กว่า, ADP)	อ้อย
ADJ+PART	(เล็ก, ADJ), (ไป, PART)	เล็ก
NOUN+AUX	(ร้าน, NOUN), (อ้อย, AUX)	ร้าน

ตารางที่ 3.11 กฎไวยากรณ์การสกัดคุณลักษณะผลิตภัณฑ์เชิงโครงสร้าง

หลังจากผ่านขั้นตอนการวิเคราะห์เชิงโครงสร้างด้วยกฎไวยากรณ์ที่กำหนดขึ้นจากบทวิจารณ์ ผู้วิจัยได้นำคำคุณลักษณะที่ได้จากประโยค มาทำการจำแนกด้วยเซตคำพ้องความหมาย (WordNet) เพื่อทำการค้นหาคำศัพท์ตัวแทนประจำเซตคำพ้องความหมายที่เกิดขึ้นในประโยค โดยลักษณะการทำงานจะเหมือนดังขั้นตอนการสกัดคุณลักษณะเชิงความถี่

### 3.7 การสกัดคุณลักษณะของผลิตภัณฑ์เชิงความถี่ร่วมกับโครงสร้าง

สำหรับขั้นตอนการสกัดคุณลักษณะเชิงความถี่ร่วมกับโครงสร้างนั้น [15] จะพิจารณาจากคำคุณลักษณะที่เกิดขึ้นในเชิงความถี่ และโครงสร้างร่วมกัน โดยคัดเลือกจากค่า TF-ICF ที่ได้จากการคำนวณ และค่าความถี่ของจำนวนความคิดเห็นที่ปรากฏคุณลักษณะเทียบกับจำนวนความคิดเห็นทั้งหมด ทั้งนี้ผู้วิจัยได้มีการกำหนดเกณฑ์การคัดเลือก (threshold) สำหรับคุณลักษณะเชิงโครงสร้าง เพื่อใช้ในการคัดเลือกคำคุณลักษณะที่มีความเหมาะสม ตัวอย่างการกำหนด threshold และการคัดเลือกคำคุณลักษณะ สามารถแสดงตัวอย่างได้ดังตาราง 3.12

ตัวอย่าง กำหนดค่า threshold ให้ TF-ICT > 0.3, Frequency > 400

คำ	TF-ICF	Frequency
Restaurant	0.305	565
Agree	0.31	493

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษานานาชาติเท่านั้น ไม่อนุญาตให้นำไปใช้ในเชิงพาณิชย์  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำ	TF-ICF	Frequency
Better	0.2	125
Charge	0.29	49
Try	0.45	20

ตารางที่ 3.12 การสกัดคุณลักษณะเชิงความถี่ร่วมกับโครงสร้าง

จากตัวอย่างข้างต้น ทำให้ผู้วิจัยสามารถคัดเลือกคำคุณลักษณะที่เกิดจากเชิงความถี่ ร่วมกับโครงสร้าง ได้แก่ Restaurant และ Agree เนื่องจากทั้งสองคำนี้มีค่า TF-ICF และ Frequency ที่สูงกว่า threshold



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 4

### ผลการทดลอง และอภิปรายผล

การดำเนินงานวิจัยเพื่อการสกัดคุณลักษณะความคิดเห็นร้านอาหารที่มีต่อการใช้บริการ ผู้วิจัยได้เก็บความคิดเห็นกลุ่มร้านอาหารญี่ปุ่น และกลุ่มร้านอาหารบุฟเฟ่ต์ โดยประกอบด้วยผลการทดลอง 3 ส่วน ได้แก่ ผลการทดลองสกัดคุณลักษณะความคิดเห็นเชิงความถี่ ผลการทดลองการสกัดคุณลักษณะเชิงโครงสร้าง และผลการทดลองการสกัดคุณลักษณะเชิงความถี่ และโครงสร้าง เป็นต้น

#### 4.1 การสกัดคุณลักษณะของผลิตภัณฑ์เชิงความถี่

หลังจากผู้วิจัยได้นำความคิดเห็นกลุ่มร้านอาหารญี่ปุ่นทั้งหมด 1,057 รายการ และความคิดเห็นกลุ่มร้านอาหารบุฟเฟ่ต์ทั้งหมด 1,638 รายการ ที่มีการบันทึกจากแพลตฟอร์มออนไลน์ด้วยวิธีการ web scraping เข้าสู่ขั้นตอนการหาค่า Term frequency - Inverse Document Frequency (TF-IDF) และขั้นตอนการค้นหาคำแทนคุณลักษณะโดยใช้เซตคำพ้องความหมายภาษาไทยที่มีการจัดกลุ่มความหมายด้วยภาษาอังกฤษ โดยสามารถรองรับคำในภาษาไทย 26,242 คำ เพื่อจำแนกคำคุณลักษณะให้เป็นกลุ่มคำพ้องความหมายภาษาอังกฤษ 498 คำ ซึ่งการจำแนกคำที่เกิดขึ้นในประโยคมีการให้จำนวนคุณลักษณะที่สกัดได้ที่แตกต่างกัน โดยประโยคที่มีความยาวมาก อาจพบจำนวนคุณลักษณะมาก รวมถึงร้านอาหารที่มีจำนวนความคิดเห็นมาก ทำให้มีจำนวนคุณลักษณะมากเช่นเดียวกัน และนอกเหนือจากนี้ผู้วิจัยได้ใช้งานคลังคำศัพท์คำพ้องความหมาย ซึ่งมีผลต่อการจำแนกคำคุณลักษณะที่แสดงถึงความหมายของคำในประโยค ผลลัพธ์ของจำนวนคุณลักษณะที่เกิดขึ้นในแต่ละดาต้าเซต สามารถแสดงได้ดังตารางที่ 4.1

กลุ่มร้านอาหาร	ดาต้าเซต	จำนวนความคิดเห็น	จำนวนคุณลักษณะที่สกัดได้
ร้านอาหารญี่ปุ่น	ร้านอาหารที่ 1	126	297
	ร้านอาหารที่ 2	247	589
	ร้านอาหารที่ 3	472	610
	ร้านอาหารที่ 4	84	497
	ร้านอาหารที่ 5	128	563
ร้านอาหารบุฟเฟ่ต์	ร้านอาหารที่ 1	719	674
	ร้านอาหารที่ 2	655	612
	ร้านอาหารที่ 3	267	649

ตารางที่ 4.1 จำนวนคุณลักษณะเชิงความถี่ในแต่ละกลุ่มร้านอาหาร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากทดลองการรันทดลองอัลกอริทึมกลุ่มร้านอาหารญี่ปุ่น สามารถแสดงค่าคะแนนน้ำหนัก TF-IDF จำแนกตามกลุ่มคำพ้องความหมาย และคำต่าง ๆ ในประโยคที่เกิดขึ้นได้ด้วยตารางที่ 4.2 ซึ่งผู้วิจัยได้คัดเลือกผลลัพธ์ 10 อันดับกลุ่มคำที่เกิดขึ้น ที่มีค่าน้ำหนัก และใจความสำคัญมากที่สุดเรียงตามลำดับ โดยบางคำศัพท์ที่ปรากฏบ่อยในประโยคอาจไม่สามารถแสดงถึงคุณลักษณะที่ทำให้ความหมายได้ดี จึงอาจต้องมีการเตรียมข้อมูลเพิ่มเติม หรือใช้คลังคำศัพท์ที่มีคุณภาพดีมากยิ่งขึ้นในการทดลอง จึงจะทำให้การตรวจพบคำในประโยคมีค่าคุณลักษณะเฉพาะทางด้านร้านอาหาร ซึ่งในปัจจุบันคลังคำศัพท์ทางด้านภาษาไทยยังพบน้อย และไม่สามารถเจาะจงเฉพาะจุดประสงค์ได้จากตารางที่ 4.2 สามารถสรุปใจความด้วยคุณลักษณะที่เกิดขึ้นในบทวิจารณ์ร้านอาหารญี่ปุ่นจากคำศัพท์ที่มีความหมายได้ดังนี้ ลูกคามีการกล่าวถึงคำว่า “สถานการณ์” เป็นจำนวนมาก เนื่องด้วยสถานการณ์โควิดทำให้ร้านอาหารลูกค้าไม่สามารถเข้าไปนั่งทานที่ร้านได้ ทางร้านจึงควรต้องทำแคมเปญในการจัดส่งอาหารเดลิเวอรี่ เพื่อรักษาลูกค้าที่ยังคงติดใช้การบริโภคอาหารของที่ร้าน “ปริมาณ” บางเมนูของร้านอาหารมีราคาแพง และมีปริมาณอาหารค่อนข้างไม่เยอะ ทำให้ร้านอาหารอาจต้องปรับปรุงเรื่องการบริหารกำไร หรือต้นทุนทางด้านปริมาณ และราคาอาหาร “เดิน” ร้านอาหารไม่อยู่ติดบริเวณริมถนน ทำให้ลูกค้าอาจต้องเดินเข้าไปบริเวณร้าน จึงอาจต้องมีการติดป้ายประชาสัมพันธ์ร้านอาหารอยู่เสมอ และจัดหาที่จอดรถที่สะดวก จึงอาจจะทำให้ลูกค้ามีความพึงพอใจมากยิ่งขึ้น จากคำคุณลักษณะสุดท้าย “มาตรฐาน” ลูกค้าผู้ใช้บริการได้มีการกล่าวชื่นชมยอมรับในมาตรฐานของร้านภายในบทวิจารณ์ ทั้งในด้านของวัตถุดิบ และการบริการของพนักงานภายในร้านอาหาร อย่างไรก็ตาม คำอื่น ๆ ที่เกิดขึ้นในประโยคแต่ไม่สามารถตีความได้ อาจต้องมีการเพิ่มคำศัพท์ในคลังคำศัพท์ ถ้าหากเป็นคำที่มีความจำเป็น เช่น คำว่า “หม้อ” ผู้ใช้บริการอาจมีความต้องการรับประทานเมนูที่ใช้หม้อในการประกอบอาหาร เป็นต้น ส่วนคำอื่น ๆ ที่ไม่มีความจำเป็น ต้องมีการคัดกรองออกจากคลังคำศัพท์คำพ้องความหมาย เพื่อให้คุณลักษณะสามารถแสดงออกได้ถึงความต้องการของผู้ใช้บริการมากยิ่งขึ้น

หมวดหมู่คำคุณลักษณะ	คุณลักษณะในบทวิจารณ์	ค่าน้ำหนัก TF-IDF
Procedure	วิธีการ	4.839452312593311
Situation	สถานการณ์	4.839452312593311
Amount	ปริมาณ	4.433987204485146
Walk	เดิน	3.740840023925201
Impress	ประทับใจ	3.2300144001592104
Pot	หม้อ	2.4197261562966554
Choose	คัด	1.4467710817689987
Manage	ดูแล	1.3821017106777884
Tell	พูด	1.1084968011212866
Quality	มาตรฐาน	0.9678904625186622

ตารางที่ 4.2 การจัดกลุ่มคุณลักษณะเชิงความถี่ร้านอาหารญี่ปุ่น

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ การใช้งานเพื่อสิ่งอื่นที่ไม่ใช่การวิจัยทางวิชาการโดยไม่ได้รับอนุญาตจะถือว่าผิดกฎหมาย

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากทดลองการรันทดลองอัลกอริทึมกลุ่มร้านอาหารบุฟเฟ่ต์ สามารถแสดงค่าคะแนนน้ำหนัก TF-IDF เฉลี่ยจำแนกตามกลุ่มคำพ้องความหมาย และคำในประโยคที่เกิดขึ้นได้ด้วยตารางที่ 4.3 ซึ่งผู้วิจัยได้คัดเลือกผลลัพธ์ 10 อันดับกลุ่มคำที่เกิดขึ้น ที่มีค่าน้ำหนัก และใจความสำคัญมากที่สุดเรียงตามลำดับอย่างเช่นตารางก่อนหน้าข้างต้นที่ปรากฏในกลุ่มร้านอาหารญี่ปุ่น ซึ่งพบว่ามีความพ้องความหมายเกิดขึ้นแตกต่างจากกลุ่มร้านอาหารญี่ปุ่นเป็นอย่างมาก จากตารางที่ 4.3 สามารถสรุปใจความด้วยคุณลักษณะที่เกิดขึ้นในบทวิจารณ์ร้านอาหารบุฟเฟ่ต์จากคำศัพท์ที่มีความหมายได้ดังนี้ ลูกค้ายมีการกล่าวถึงคำว่า “การพัฒนา” เป็นจำนวนมาก จากการสรุปใจความสำคัญทำให้ทราบว่าลูกค้ายมีความพึงพอใจที่ร้านอาหารมีการพัฒนาคุณภาพอยู่เสมอจากการเข้าใช้บริการในแต่ละครั้ง “การปรับปรุง” จากบทวิจารณ์พบว่าลูกค้ายได้เห็นถึงการปรับปรุงคุณภาพการให้บริการควบคู่กับการพัฒนาอยู่เสมอ “ระยะเวลา” ลูกค้ายมีการกล่าวถึงระยะเวลาจำกัดในการรับประทานอาหารภายในร้านบุฟเฟ่ต์ ร้านอาหารจึงอาจต้องมีการขุดเจาะด้านคุณภาพของอาหาร และความคุ้มค่าในระยะเวลาที่จำกัด เพื่อให้ลูกค้ายมีความพึงพอใจที่มากขึ้น ในขณะที่เดียวกันลูกค้ายมีการวิจารณ์เกี่ยวกับ “ความคุ้มค่า” เช่นเดียวกัน ร้านอาหารจึงต้องรักษาจุดเด่นนี้ เพื่อให้ลูกค้ายมีความพึงพอใจในการใช้บริการครั้งต่อไป จากคำคุณลักษณะ “ทางเลือก” ร้านอาหารบุฟเฟ่ต์ยังคงเป็นทางเลือกของผู้ใช้บริการบางส่วน เนื่องจากผู้ให้บริการบางท่านคำนึงถึงสถานที่ จึงอาจต้องมีการพิจารณาทำเลในการตั้งร้านอาหาร หรือหาช่องทางในการเปิดสาขาใหม่ เพื่อรองรับกลุ่มลูกค้ายที่เพิ่มมากขึ้น จากจำนวนบทวิจารณ์ร้านอาหารบุฟเฟ่ต์ที่มากขึ้น เมื่อเทียบกับบทวิจารณ์ร้านอาหารญี่ปุ่น พบว่าอาจมีคำคุณลักษณะที่สมเหตุสมผลมากขึ้น แต่บางคำอาจเกินความจำเป็น เนื่องจากมีความหมายของคำคุณลักษณะใกล้เคียงกัน และมีค่า TF-IDF ใกล้เคียงกัน บางคำจึงอาจยุบรวมกันได้ในคลังคำศัพท์คำพ้องความหมาย โดยเฉพาะคำว่า “การพัฒนา” และ “การปรับปรุง”

หมวดหมู่คำคุณลักษณะ	คุณลักษณะในบทวิจารณ์	ค่าน้ำหนัก TF-IDF
Flaw	ข้อด้อย	6.4680601411351315
Development	การพัฒนา	6.4680601411351315
Improvement	การปรับปรุง	6.062595033026967
Period of time	ระยะเวลา	5.2152971726397634
Dignity	ความคุ้มค่า	5.081765780015241
Selection	ตัวเลือก	5.081765780015241
Attention	ความตั้งใจ	3.2340300705675658
Option	ทางเลือก	3.0312975165134834
Trend	กระแส	2.6847239262335107
Request	ขอเรียกร้อง	0.8085075176418914

ตารางที่ 4.3 การจัดกลุ่มคุณลักษณะเชิงความถี่ร้านอาหารบุฟเฟ่ต์

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์หรือการเรียงในเพื่อวัตถุประสงค์เท่านั้น เมื่อผู้ยู่เห็นในข้อระเบียบข้อบังคับด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการสกัดคุณลักษณะของผลิตภัณฑ์เชิงความถี่พบว่ามีย่อยกลุ่มคำคุณลักษณะที่มีความหมายคล้ายกันปรากฏขึ้นในประโยคเป็นจำนวนมากจากการพิจารณาคำนำน้าหนัก TF-IDF สามารถพบเห็นได้จากดาต้าเซตร้านอาหารบุฟเฟ่ต์ เช่น Development - การพัฒนา และ Improvement - การปรับปรุง เป็นต้น จึงอาจต้องมีการพัฒนาคำศัพท์คำพ้องความหมายเพิ่มเติม และพิจารณาเชิงโครงสร้าง เพื่อให้มีลักษณะการจำแนกกลุ่มที่ดียิ่งขึ้น และลดความซ้ำซ้อนของกลุ่ม โดยคำคุณลักษณะบางคำสามารถรวบรวมคำเพื่อจัดกลุ่มเดียวกันได้ ตัวอย่างประโยคแสดงได้ดังรูปที่ 4.1 และ 4.2

อาหารมีน้อยไป เป็นร้านที่ต้องมีการพัฒนาต่อไปอีก

รูปที่ 4.1 ตัวอย่างประโยคที่พบคำคุณลักษณะเชิงความถี่ “พัฒนา”

เราน่าจะปรับปรุงใหม่ให้รสชาติเด่นกว่านี้เพราะเวลานำกุ้งหรืออาหารซีฟู้ดมาจิ้มมันไม่ได้ช่วยชูรสชาติหรือทำให้อยากกินต่อ

รูปที่ 4.2 ตัวอย่างประโยคที่พบคำคุณลักษณะเชิงความถี่ “ปรับปรุง”

ถ้าหากมีประโยคดังรูปที่ 4.1 และ 4.2 ประกอบด้วยคำที่มีความหมายเหมือนกันจำนวนมากในบทวิจารณ์ จะทำให้เกิดหมวดหมู่คำคุณลักษณะที่มีค่าน้ำหนัก TF-IDF ซ้ำซ้อนใกล้เคียงกันจำนวนที่มากกว่าหนึ่งหมวดหมู่ และเนื่องจากโครงสร้างที่ซับซ้อนในการพิจารณาคำในเชิงความถี่ การสกัดคุณลักษณะของคำบางชนิดอาจมีรูปแบบอิสระ ไม่จำเป็นต้องการประโยคที่กล่าวถึงเพื่อสื่อความหมายก่อนหน้า [17] การสกัดคุณลักษณะเชิงโครงสร้างจึงเข้ามาช่วยให้การทำงานมีประสิทธิภาพยิ่งขึ้น นอกเหนือจากการใช้ความถี่ และพิจารณาคำนำน้าหนัก [15]

## 4.2 การสกัดคุณลักษณะของผลิตภัณฑ์เชิงโครงสร้าง

ผู้วิจัยได้นำข้อมูลกลุ่มร้านอาหารญี่ปุ่น และกลุ่มร้านอาหารบุฟเฟ่ต์ตั้งเช่นกระบวนการข้างต้นเข้าสู่ขั้นตอนการสกัดคุณลักษณะของผลิตภัณฑ์เชิงโครงสร้าง โดยมีการทดลองกำหนดกฎโครงสร้างไวยากรณ์ในการสกัดคุณลักษณะตามลักษณะของข้อมูลเบื้องต้นที่สามารถเกิดขึ้นได้ในบทวิจารณ์ กฎโครงสร้างไวยากรณ์นี้ได้ผ่านการคัดกรองคุณลักษณะ โดยใช้เซตคำพ้องความหมายภาษาไทยที่มีการจัดกลุ่มความหมายด้วยภาษาอังกฤษ เนื่องจากมีกฎไวยากรณ์ต่างๆเกิดขึ้นมากมาย จึงสามารถแสดงตัวอย่างกฎไวยากรณ์เบื้องต้นได้ดังตารางที่ 4.4 สามารถตั้งกฎไวยากรณ์ได้จากกรณีต่าง ๆ ของคำกรณิคำที่เกิดขึ้น การวิเคราะห์โครงสร้างในภาษาไทยประกอบด้วยชนิดของวลีจำนวน 4 ส่วน คือนามวลี กริยาวลี วิเศษณ์วลี สถานวลี กาลวลี [17] จากการค้นคว้างานวิจัยที่เกี่ยวข้อง ทำให้ผู้วิจัยได้ใช้นามวลีในการตั้งกฎไวยากรณ์เนื่องจากนามวลี (Noun phrase) สามารถแสดง ความสำคัญระหว่างคุณลักษณะผลิตภัณฑ์ และข้อคิดเห็นได้เป็นอย่างดี [15] รวมทั้งได้มีการตั้งกฎไวยากรณ์โครงสร้างทางภาษาที่เกิดขึ้นจริง (Corpus-based) ต้องมีการใช้งานความรู้ที่ได้จากคลังข้อมูลการใช้ภาษาที่เกิดขึ้นจริงในประโยคที่อาจไม่เป็นไปตามไวยากรณ์ภาษาศาสตร์ (ร้าน, NOUN), (ปิ้งย่าง, NOUN)

ผู้วิจัยได้คัดเลือกคำว่า “(ปิ้งย่าง, NOUN)” เนื่องจากสามารถแสดงคุณลักษณะได้ดีกว่า และอาจพบ

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ได้น้อยกว่าคำว่า “ร้าน” ซึ่งการสกัดคุณลักษณะจากคำที่มีจำนวนมากอาจทำให้คุณลักษณะสื่อถึงภาพรวมเกินความจำเป็น (คนญี่ปุ่น, NOUN), (เลี้ยง, VERB) ผู้วิจัยได้คัดเลือกคำว่า “(คนญี่ปุ่น, NOUN)” เนื่องจากประธานของประโยค (NOUN) สามารถสื่อได้ถึงคุณลักษณะหลักของประโยคได้นอกเหนือจากนี้ยังสามารถนำไปค้นหาคำแสดงความรู้สึกที่เชื่อมโยงจากประธานได้ในอนาคต นอกเหนือจากนี้ยังสามารถนำไปค้นหาคำแสดงความรู้สึกที่เชื่อมโยงจากประธานได้ในอนาคต กฎไวยากรณ์ร่วม NOUN พบในคำ bigram อื่น ๆ เพิ่มเติม สามารถตั้งกฎในลักษณะนามวลี ได้แก่ NOUN + ADV, NOUN + ADJ, NOUN + ADP, NOUN + DET และ AUX + NOUN เป็นต้น ส่วนคำคุณลักษณะ (ต้อง, AUX), (จอง, VERB) ผู้วิจัยได้คัดเลือก “(จอง, VERB)” เนื่องจากกิริยานุเคราะห์ (AUX: auxiliary verb) เป็นกริยาช่วย คำว่า “(จอง, VERB)” จึงสามารถแสดงคุณลักษณะที่มีความหมายได้ดีกว่า ร้านอาหารสามารถเข้าใจได้ว่า ลูกค้านำกล่าวถึงการจองได้อย่างชัดเจน คำคุณลักษณะต่อมา (อ้อย, ADV), (กว่า, ADP) ผู้วิจัยได้คัดเลือก “(อ้อย, ADV)” เนื่องจากคำบุพพท (ADP: adposition, preposition) เป็นคำเชื่อมสรรพนาม อาจต้องการคำอื่นเพื่อสื่อความหมายเพิ่ม คำคุณลักษณะสุดท้ายจากการคัดกรองด้วยกฎไวยากรณ์ (เล็ก, ADJ), (ไป, PART) ผู้วิจัยได้คัดเลือก (เล็ก, ADJ) เนื่องจากคำเชื่อม (PART: particle) ต้องประกอบด้วยคำอื่นสื่อความหมายเพิ่มเติม ส่วนผลลัพธ์ของจำนวนคุณลักษณะที่เกิดขึ้นในแต่ละดาต้าเซต สามารถแสดงได้ดังตารางที่ 4.5

กฎไวยากรณ์	คำที่เกิดขึ้น	คำคุณลักษณะ
NOUN+NOUN	(ร้าน, NOUN), (ปิ้งย่าง, NOUN)	(ปิ้งย่าง, NOUN)
NOUN+VERB	(คนญี่ปุ่น, NOUN), (เลี้ยง, VERB)	(คนญี่ปุ่น, NOUN)
NOUN+ADV	(บริการ, NOUN), (เร็ว, ADV)	(บริการ, NOUN)
NOUN+ADJ	(เนื้อ, NOUN), (ธรรมดา, ADJ)	(เนื้อ, NOUN)
NOUN+ADP	(บริการ, NOUN), (ทั่วถึง, ADP)	(บริการ, NOUN)
NOUN+DET	(ราคา, NOUN), (นี้, DET)	(ราคา, NOUN)
AUX+NOUN	(เป็น, AUX), (ญี่ปุ่น, NOUN)	(ญี่ปุ่น, NOUN)
AUX+VERB	(ต้อง, AUX), (จอง, VERB)	(จอง, VERB)
ADV+ADP	(อ้อย, ADV), (กว่า, ADP)	(อ้อย, ADV)
ADJ+PART	(เล็ก, ADJ), (ไป, PART)	(เล็ก, ADJ)

ตารางที่ 4.4 กฎไวยากรณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กลุ่มร้านอาหาร	ดาต้าเซต	จำนวนความคิดเห็น	จำนวนคุณลักษณะที่สกัดได้
ร้านอาหารญี่ปุ่น	ร้านอาหารที่ 1	126	409
	ร้านอาหารที่ 2	247	517
	ร้านอาหารที่ 3	472	540
	ร้านอาหารที่ 4	84	394
	ร้านอาหารที่ 5	128	492
ร้านอาหารบุฟเฟ่ต์	ร้านอาหารที่ 1	719	616
	ร้านอาหารที่ 2	655	557
	ร้านอาหารที่ 3	267	563

ตารางที่ 4.5 จำนวนคุณลักษณะเชิงโครงสร้างในแต่ละกลุ่มร้านอาหาร

จากทดลองการรันทดลองอัลกอริทึมกลุ่มร้านอาหารญี่ปุ่น กลุ่มคำพ้องความหมายที่เกิดขึ้นจากการค้นหาคำในประโยค ซึ่งเกิดจากผลลัพธ์ของการดำเนินการด้วยกฎไวยากรณ์ที่ผู้วิจัยได้กำหนดไว้ดังตารางที่ 4.4 และนำคำในประโยคที่ถูกจัดกลุ่มด้วยคำพ้องความหมายมาคิดค่า TF-IDF เพื่อคัดเลือกคำตัวแทนในประโยคที่มีใจความสำคัญมากที่สุดอันดับหนึ่งในแต่ละคำพ้อง พร้อมทั้งเฉลี่ยค่า TF-IDF ประจាកุ่มคำพ้องใหม่ สามารถแสดงผลลัพธ์กลุ่มคำพ้องความหมายที่ผ่านการสกัดคุณลักษณะเชิงโครงสร้าง และผ่านการจัดกลุ่มคำพ้องความหมายด้วยคลังคำศัพท์แสดงได้ดังตารางที่ 4.6 ซึ่งผู้วิจัยได้คัดเลือกผลลัพธ์ 10 อันดับกลุ่มคำที่เกิดขึ้น ที่มีค่าน้ำหนัก และใจความสำคัญมากที่สุดเรียงตามลำดับ จากตารางที่ 4.6 สามารถสรุปใจความด้วยคุณลักษณะที่เกิดขึ้นในบทวิจารณ์ร้านอาหารญี่ปุ่นจากคำศัพท์ที่มีความหมายได้ดังนี้ ลูกค้านี้มีการกล่าวถึงคำว่า “ปริมาณ” จากบทวิจารณ์ที่เกิดขึ้น ร้านอาหารสามารถทราบได้ว่าลูกค้ามีความสนใจเกี่ยวกับปริมาณอาหารของแต่ละเมนู “รส” ลูกค้ามีการให้คำชมเชยเกี่ยวกับรสชาติอาหาร ผู้ประกอบการจึงควรรักษามาตรฐานของรสชาติให้มีความสม่ำเสมอ เพื่อให้กลุ่มลูกค้ายังคงไว้วางใจในการรับประทานอาหารที่ร้าน “หน้าตา” พนักงานของร้านควรมีหน้าตาที่ยิ้มแย้ม เต็มใจในการให้บริการ คอยเอาใจใส่ลูกค้าเป็นอย่างดี ส่วนคุณลักษณะสุดท้ายมีคำว่า “มาตรฐาน” อาจกล่าวได้ว่าร้านอาหารได้รับการยอมรับจากลูกค้าในการเป็นที่ยอมรับ ทางด้านวัตถุดิบ และการบริการ

หมวดหมู่คำคุณลักษณะ	คุณลักษณะในบทวิจารณ์	ค่าน้ำหนัก TF-IDF
Think of	นึกถึง	4.839452312593311
Give tongue to	พูด	4.433987204485146
Countenance	ยอมรับ	4.433987204485146
Amount	ปริมาณ	4.433987204485146
Request	ขอ	3.9231615807191558

เอกสารนี้เป็นเอกสารสำหรับการใช้งานเพื่อการศึกษานี้เท่านั้น ไม่อนุญาตให้มีการคัดลอกหรือเผยแพร่โดยไม่ได้รับอนุญาต

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หมวดหมู่คำคุณลักษณะ	คุณลักษณะในบทวิจารณ์	ค่าน้ำหนัก TF-IDF
Sensation	รส	3.740840023925201
Provide	หา	3.586689344097943
Impress	ประทับใจ	3.2300144001592104
Appearance	หน้าตา	1.613150770864437
Quality	มาตรฐาน	0.9678904625186622

ตารางที่ 4.6 การจัดกลุ่มคุณลักษณะเชิงโครงสร้างร้านอาหารญี่ปุ่น

จากทดลองการรันทดลองอัลกอริทึมกลุ่มร้านอาหารบุฟเฟ่ต์ พบว่ามีคำคุณลักษณะบางคำสื่อความหมายที่เข้าใจมากกว่ากลุ่มร้านอาหารญี่ปุ่นที่มีจำนวนบทวิจารณ์น้อยกว่า ซึ่งผู้วิจัยได้เห็นว่าถ้าหากมีข้อมูลบทวิจารณ์ที่เยอะขึ้น ประกอบกับคลังคำศัพท์ที่สมบูรณ์จะยิ่งทำให้ได้กลุ่มคำพ้องความหมายสามารถสื่อใจความได้ชัดชัด โดยตัวอย่างผลลัพธ์การทดลองสามารถแสดงได้ดังตารางที่ 4.7 สามารถสรุปใจความด้วยคุณลักษณะที่เกิดขึ้นในบทวิจารณ์ร้านอาหารบุฟเฟ่ต์จากคำศัพท์ที่มีความหมายได้ดังนี้ ลูกค้านี้มีการกล่าวถึงคำว่า “การพัฒนา” จากบทวิจารณ์ที่เกิดขึ้น ซึ่งเกิดขึ้นเช่นเดียวกันกับตารางที่ 4.3 ลูกค้ามีความพึงพอใจที่ร้านมีการพัฒนาคุณภาพอยู่เสมอ “ส่วนลด” ลูกค้าให้ความสนใจเกี่ยวกับส่วนลดค่าบริการที่ทางร้านอาหารกำหนดให้แก่ผู้สมัครสมาชิก “ฝูงชน” ลูกค้าพึงพอใจเกี่ยวกับการจัดคิวผู้เข้ารับบริการ และประชาสัมพันธ์การเรียกคิวเป็นอย่างดี “ความโกลาหล” ลูกค้ามีการกล่าวถึงการให้บริการที่มีลูกค้าอยู่หน้าร้านจำนวนมาก ซึ่งมีความสัมพันธ์กับคำว่า “ฝูงชน” ดังข้างต้น อาจต้องมีการพัฒนาคลังคำศัพท์เพิ่มขึ้น เพื่อลดความซ้ำซ้อนของคำ และกำหนดกฎไวยากรณ์เพิ่มขึ้น เพื่อขจัดคำบางคำที่จับใจความได้ยาก

หมวดหมู่คำคุณลักษณะ	คุณลักษณะในบทวิจารณ์	ค่าน้ำหนัก TF-IDF
Speech	คำพูด	6.4680601411351315
Development	การพัฒนา	6.4680601411351315
Croak	บ่น	5.774912960575186
Reduction	ส่วนลด	4.116684883971653
Battle	แข่งขัน	3.2340300705675658
Option	ทางเลือก	3.0312975165134834
Flock	ฝูงชน	2.1560200470450437
Disorder	ความโกลาหล	2.1560200470450437
Worth	มูลค่า	1.738432390879921
Receive	ยอมรับ	1.2409956860897142

ตารางที่ 4.7 การจัดกลุ่มคุณลักษณะเชิงความโครงสร้างร้านอาหารบุฟเฟ่ต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการสกัดคุณลักษณะของผลิตภัณฑ์เชิงโครงสร้างพบว่าหมวดหมู่คำคุณลักษณะที่มีความหมายคล้ายกันไม่มีการปรากฏขึ้นเหมือนกับวิธีการเชิงความถี่ก่อนหน้านี้ เนื่องจากมีการพิจารณาคำในประโยคแต่ละชนิดด้วยชนิดของคำที่ต่างกัน จึงมีความหลากหลายของคำที่มากกว่า และด้วยความหลากหลายทางธรรมชาติของภาษาเหล่านี้ที่เกิดขึ้นทำให้เราไม่สามารถสร้างกฎไวยากรณ์ที่ครอบคลุมได้ทั้งหมดในบทวิจารณ์ สามารถแสดงตัวอย่างประโยคที่ประกอบด้วยคำคุณลักษณะนามวลีได้จากดาต้าเซตร้านอาหารบุฟเฟต์ เช่น Development - การพัฒนา และ Reduction - ส่วนลด เป็นต้น ได้ดังรูปที่ 4.3 และ 4.4

อาหารมีน้อยไป เป็นร้านที่ต้องมีการพัฒนาต่อไปอีก

รูปที่ 4.3 ตัวอย่างประโยคที่พบคำคุณลักษณะเชิงโครงสร้าง “การพัฒนา”

ต้องการเลือกรับเป็นส่วนลด ที่ได้จากการสมัครสมาชิก มีความคุ้มค่าในราคาที่เหมาะสม

รูปที่ 4.4 ตัวอย่างประโยคที่พบคำคุณลักษณะเชิงโครงสร้าง “ส่วนลด”

#### 4.3 การสกัดคุณลักษณะของผลิตภัณฑ์เชิงความถี่ร่วมกับโครงสร้าง

การรวบรวมคุณลักษณะที่เกิดขึ้นทั้งในเชิงความถี่ร่วมกับเชิงโครงสร้าง ผู้วิจัยได้ทำการเริ่มต้นการทดลองจากการพิจารณาคำคุณลักษณะที่มีค่า TF-IDF น้อยที่สุด และค่าความถี่ของจำนวนความคิดเห็นที่ ปรากฏคุณลักษณะเทียบกับจำนวนความคิดเห็นทั้งหมดน้อยที่สุด (Occurrence Frequency: OF) จากหัวข้อที่ 4.1 และ 4.2 ดาต้าเซตร้านอาหารบุฟเฟต์ที่มีบทวิจารณ์จำนวนมาก และสามารถสื่อใจความได้ดี โดยกำหนด threshold ด้วยค่า TF-IDF > 1.2 เนื่องจากผลลัพธ์ตัวอย่างของตารางที่ 4.7 คำพ้องความหมายที่สื่อถึงคุณลักษณะ “Receive-ยอมรับ” มีค่า TF-IDF ที่ต่ำที่สุด 1.2 และจากการคำนวณค่าความถี่ของประโยคที่พบคำนี้ในบทวิจารณ์เทียบกับจำนวนบทวิจารณ์ทั้งหมด (Occurrence Frequency) มีค่าเท่ากับ 6/473 หรือประมาณ 1% จึงมีการกำหนด threshold ด้วยค่า Occurrence Frequency > 1% อีกจำนวนหนึ่งจำนวนคุณลักษณะแสดงได้ดังตารางที่ 4.8

กลุ่มร้านอาหาร	ดาต้าเซต	จำนวนความคิดเห็น	จำนวนคุณลักษณะที่สกัดได้
ร้านอาหารญี่ปุ่น	ร้านอาหารที่ 1	126	297
	ร้านอาหารที่ 2	247	410
	ร้านอาหารที่ 3	472	463
	ร้านอาหารที่ 4	84	281
	ร้านอาหารที่ 5	128	405
ร้านอาหารบุฟเฟต์	ร้านอาหารที่ 1	719	533
	ร้านอาหารที่ 2	655	455
	ร้านอาหารที่ 3	267	465

ตารางที่ 4.8 จำนวนคุณลักษณะเชิงความถี่ร่วมกับโครงสร้างในแต่ละกลุ่มร้าน  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากทดลองการรันทดลองอัลกอริทึมกลุ่มร้านอาหารญี่ปุ่น พบว่าคำคุณลักษณะบางชนิดที่ เคยปรากฏในขั้นตอนการสกัดคุณลักษณะเชิงความถี่ และการสกัดคุณลักษณะเชิงโครงสร้างก่อนหน้า ได้หายไป ทำให้ขาดคุณลักษณะที่สื่อถึงความหมายบางประโยคได้น้อยลง สามารถแสดงได้ดังตารางที่ 4.9 สามารถสรุปใจความด้วยคุณลักษณะที่เกิดขึ้นในบทวิจารณ์ร้านอาหารญี่ปุ่นจากคำศัพท์ที่มีความหมายได้ดังนี้ ลูกค้ำมีการกล่าวถึงคำว่า “คุ่ม” จากบทวิจารณ์ที่เกิดขึ้น ผู้วิจัยสามารถวิเคราะห์ จากบทวิจารณ์ได้ว่าถึงแม้ร้านอาหารจะทำตลาดในราคาค่อนข้างแพง แต่ก็ยังมีความคุ้มค่า ลูกค้ำให้ การยอมรับ ส่วนคำว่า “หักคะแนน” ลูกค้ำมีการกล่าวถึงการหักคะแนนอุปกรณ์ประกอบอาหาร แต่ อาจจะไม่ได้ลงรายละเอียดถึงความสะดวก หรือคุณภาพการให้บริการ “แนะนำ” ลูกค้ำมีเขียนบท วิจารณ์ทางด้านการแนะนำบริการจำนวนหนึ่ง “ปริมาณ” ลูกค้ำได้มีการเขียนบทวิจารณ์บอกถึง ปริมาณอาหารที่ได้ในเมนู เพื่อแนะนำผู้บริโภคท่านอื่น ส่วนคำคุณลักษณะ “หน้าตา” พบในในตาราง ที่ 4.6 เช่นเดียวกัน บทวิจารณ์กล่าวได้ว่าพนักงานของร้านค้ำมีหน้าตาที่ยิ้มแย้ม เต็มใจในการ ให้บริการ คอยเอาใจใส่ลูกค้ำเป็นอย่างดี อาจกล่าวได้ตรงกันว่าร้านอาหารเป็นที่ยอมรับ ทางด้าน วัตถุประสงค์ และการบริการ

หมวดหมู่คำคุณลักษณะ	คุณลักษณะในบทวิจารณ์	ค่าน้ำหนัก TF-IDF
Think of	นึกถึง	4.839452312593311
Give tongue to	พูด	4.433987204485146
Cover	คุ่ม	3.740840023925201
Take away	หักคะแนน	4.839452312593311
Clean	ถู	2.4197261562966554
Induce	แนะนำ	2.3137236682850553
Quantity	ปริมาณ	2.216993602242573
Appearance	หน้าตา	1.613150770864437
Room	เนื้อที่	1.34969307454852
Eat	รับประทาน	1.2098630781483277

ตารางที่ 4.9 การจัดกลุ่มคุณลักษณะเชิงความถี่ร่วมกับโครงสร้างร้านอาหารญี่ปุ่น

จากทดลองการรันทดลองอัลกอริทึมกลุ่มร้านอาหารบุฟเฟ่ต์ พบว่าคำคุณลักษณะบางส่วนที่ สำคัญได้ผลลัพธ์ที่ลดลง แต่สามารถสื่อความหมาย และจับใจความได้ โดยสามารถแสดงผลที่ได้ใน ตารางที่ 4.10 สามารถสรุปใจความด้วยคุณลักษณะที่เกิดขึ้นในบทวิจารณ์ร้านอาหารบุฟเฟ่ต์จาก คำศัพท์ที่มีความหมายได้ดังนี้ ลูกค้ำมีการกล่าวถึงคำว่า “ความผิดหวัง” จากบทวิจารณ์ที่เกิดขึ้น ลูกค้ำส่วนใหญ่ยังได้มีการวิจารณ์ในเชิงก้ำกวม สามารถตีความได้ว่าลูกค้ำไม่ได้รับความผิดหวังในการ ใช้บริการ หรืออาจตีความได้ว่ามีความพึงพอใจในการใช้บริการนั้นเอง ส่วนคำว่า “ภาษี” ภายในบท

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ในพจนานุกรมภาษาไทย-อังกฤษ โดยผู้จัดทำขึ้นเพื่อประโยชน์ในการศึกษา ค้นคว้าวิจัยเท่านั้น ไม่สามารถนำ ไปใช้  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิจารณ์พบว่า ร้านค้ามีการคิดบางรายการที่รวมภาษีแล้ว จึงอาจต้องคำนึงถึงการคำนวณรายการค่าบริการแบบแยกภาษี เพื่อให้เกิดความสบายใจมากยิ่งขึ้น “ระยะเวลา” ภายในบทวิจารณ์ผู้ประกอบการสามารถทราบได้ว่าลูกค้ามีความสนใจเรื่องระยะเวลาในการรับประทานอาหารเช้า และระยะเวลาในการเสิร์ฟเมนูอาหาร ผู้ให้บริการสามารถหาช่องทางในการเพิ่มความพึงพอใจแก่ลูกค้าในด้านความรวดเร็วในการให้บริการ ปรับระยะเวลาในการรับประทานอาหารเช้าได้ “ทำความสะอาด” ภายในบทวิจารณ์ลูกค้าได้มีการวิจารณ์เกี่ยวกับความสะอาดในการทำความสะอาดบริเวณร้าน อาทิ เช่น โต๊ะ เก้าอี้ และอุปกรณ์ต่าง ๆ โดยรอบ ร้านอาหารจึงควรรักษามาตรฐานคงไว้ เพื่อรักษารฐานลูกค้าที่มีความชื่นชอบเรื่องความสะอาด

หมวดหมู่คำคุณลักษณะ	คุณลักษณะในบทวิจารณ์	ค่าน้ำหนัก TF-IDF
Defeat	ความผิดหวัง	6.4680601411351315
Competition	การแข่งขัน	6.4680601411351315
Duty	ภาษี	6.4680601411351315
Croak	บ่น	5.774912960575186
Time period	ระยะเวลา	5.2152971726397634
Sight	วิว	3.2340300705675658
Scavenge	ทำความสะอาด	3.0312975165134834
Disposition	อัธยาศัย	1.5156487582567417
Surpass	ข้าม	1.3038242931599409
Magnitude	มิติ	1.2936120282270263

ตารางที่ 4.10 การจัดกลุ่มคุณลักษณะเชิงความถี่ร่วมกับโครงสร้างร้านอาหารบุฟเฟ่ต์

จากการสกัดคุณลักษณะของผลิตภัณฑ์เชิงความถี่ร่วมกับโครงสร้างหมวดหมู่คุณลักษณะที่มีความหมายคล้ายกันไม่มีการปรากฏขึ้นเหมือนกับวิธีการเชิงความถี่ก่อนหน้านี้เช่นเดียวกัน แต่อาจต้องมีการปรับค่า threshold ให้เหมาะสมแก่การใช้งาน เพื่อลดค่าที่ไม่สอดคล้องที่เกิดขึ้นในประโยคที่เกิดจากการรวบรวมค่าตอบร่วมกันระหว่างคุณลักษณะเชิงความถี่ และโครงสร้าง สามารถแสดงตัวอย่างประโยคที่ประกอบด้วยคำคุณลักษณะนามวลีได้จากดาด้าเซตร้านอาหารบุฟเฟ่ต์ เช่น Duty - ภาษี และ Time period - ระยะเวลา เป็นต้น ดังรูปที่ 4.5 และ 4.6

ราคาอาหารแพงมาก มีการคิดรวมค่าบริการภาษี ถ้าไม่มีการชาร์ตภาษีแพงคงถูกกว่านี้

รูปที่ 4.5 ตัวอย่างประโยคที่พบคำคุณลักษณะเชิงความถี่ร่วมโครงสร้าง “ภาษี”

ต้องการเลือกรับเป็นส่วนลด ที่ได้จากการสมัครสมาชิก มีความคุ้มค่าในราคาที่พอเหมาะ

รูปที่ 4.6 ตัวอย่างประโยคที่พบคำคุณลักษณะเชิงความถี่ร่วมโครงสร้าง “ส่วนลด”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการศึกษาสกัดคุณลักษณะในประโยคด้วยอัลกอริทึมพบว่าคุณลักษณะที่ได้นั้นบางครั้งอาจมีคำที่เกิดขึ้นซ้ำได้ในประโยคหลายครั้งสำหรับประโยคที่มีความยาว หรือพบได้หลายในหลายประโยค จึงก่อให้เกิดความถี่สะสมเพิ่มขึ้นในขั้นตอนการสกัดคุณลักษณะเชิงความถี่ หมวดหมู่คำคุณลักษณะที่ใช้เป็นข้อสรุปของบทวิจารณ์จึงมีลักษณะที่ซ้ำซ้อนกันตามความถี่ของคำที่พบในบทวิจารณ์ ซึ่งแตกต่างจากเทรนของการสกัดคุณลักษณะเชิงโครงสร้าง และการสกัดคุณลักษณะเชิงความถี่ร่วมกับโครงสร้าง ที่จะให้ผลของการสกัดคุณลักษณะที่มีความหลากหลายตามธรรมชาติภาษาของข้อมูล เนื่องจากการสกัดโครงสร้างตามกฎไวยากรณ์ และมีการคัดกรองคำที่มีความน้ำหนักความถี่ และค่าความถี่ของประโยคที่พบคุณลักษณะตามที่กำหนด โดยการทดลองทั้งหมดทำให้คุณลักษณะที่ผู้วิจัยสามารถสกัดออกมา เป็นประโยชน์สูงสุดระหว่างผู้ประกอบการร้านค้า และผู้ใช้บริการ สามารถตอบสนองการบริการได้อย่างรวดเร็ว เข้าใจจุดแข็งจุดอ่อนของธุรกิจ ผู้ให้บริการสามารถทำแคมเปญส่งเสริมธุรกิจ หรือเพิ่มความพึงพอใจแก่ลูกค้าในการให้บริการ และสามารถทำการวิเคราะห์เพิ่มเติมในเชิงลึกทางด้านบวกด้านลบของคุณลักษณะเพิ่มเติมทางด้านภาษาไทยได้ในอนาคต



## บทที่ 5

# สรุปผลการศึกษา และข้อเสนอแนะ

### 5.1 สรุปผลงานวิจัย

งานวิจัยฉบับนี้ได้นำเสนออัลกอริทึมการสกัดคุณลักษณะ ซึ่งมีการสกัดคุณลักษณะเชิงความถี่ การสกัดคุณลักษณะเชิงโครงสร้าง และคุณลักษณะเชิงความถี่ร่วมกับโครงสร้าง สำหรับคุณลักษณะเชิงความถี่นั้นให้ผลลัพธ์ที่ไม่สมเหตุผล เนื่องจากมีการพิจารณาที่ความถี่ของคำเป็นหลัก ทำให้ไม่สามารถสกัดคำคุณลักษณะที่สื่อความหมายของบทวิจารณ์ได้โดยตรง แต่อาจมีค่าความถี่ และจำนวนที่เกิดขึ้นในประโยคได้น้อย นอกเหนือจากนี้ในการจัดหมวดหมู่คำคุณลักษณะด้วยคลังคำศัพท์คำพ้องความหมาย เพื่อให้ผลลัพธ์ที่สื่อความหมายนั้น ยังทำไม่ได้ดีเท่าที่ควร เนื่องจากคลังคำศัพท์ภาษาไทยในปัจจุบันนั้นยังพบน้อย และคลังคำศัพท์ให้ผลลัพธ์โดเมนที่จำกัดเฉพาะข้อมูลที่กำหนด หากเปรียบเทียบกับ การสกัดคุณลักษณะเชิงโครงสร้าง สามารถให้ผลลัพธ์ที่เจาะจงกับปัญหาได้ดีกว่า และมีค่าที่ไม่จำเป็นน้อยลง แต่ก็มีข้อจำกัดที่กฎไวยากรณ์อาจกำหนดได้ไม่ครอบคลุมกับลักษณะบทวิจารณ์ทั้งหมด รวมทั้งการจัดหมวดหมู่ด้วยคลังคำศัพท์ยังให้ผลที่ไม่เป็นไปตามลักษณะของบทวิจารณ์ ต้องใช้ความเข้าใจบริบทในการจำแนกผลลัพธ์หลายคำ จากปัญหาที่เกิดขึ้นทั้งการสกัดคุณลักษณะเชิงความถี่ คุณลักษณะเชิงโครงสร้าง ทำให้ผู้วิจัยได้ทดลองนำการสกัดคุณลักษณะเชิงความถี่ร่วมกับการสกัดคุณลักษณะเชิงโครงสร้าง ซึ่งให้ผลลัพธ์ใกล้เคียงกับการสกัดคุณลักษณะเชิงโครงสร้าง แต่สามารถปรับค่าให้เหมาะสมตามลักษณะบทวิจารณ์ด้วยการปรับ threshold ทั้งการคัดกรองคำที่เกิดขึ้นในประโยคด้วยการกำหนดขั้นต่ำค่า TF-IDF และการกำหนดขั้นต่ำค่าความถี่ที่คำคุณลักษณะเกิดขึ้นในประโยค นอกจากนี้ข้อมูลที่ผู้วิจัยนำมาใช้พัฒนาอัลกอริทึมไม่พบดาต้าเซตบทวิจารณ์ร้านอาหารแบบภาษาไทย เกี่ยวกับการสกัดคุณลักษณะของประโยค ผู้วิจัยจึงต้องเก็บข้อมูลออนไลน์จำนวนมาก โดยใช้เทคนิค web scraping ซึ่งไม่มีการ label ข้อมูล ผู้วิจัยต้องใช้ความเข้าใจผลลัพธ์จำนวนมาก เพื่อศึกษาผลลัพธ์ที่เกิดขึ้น ซึ่งแตกต่างจากการวิเคราะห์ข้อมูลภาษาอังกฤษที่มีข้อมูลตัวอย่าง พร้อมทั้งมีการกำหนด label ข้อมูล ทำให้สามารถวิเคราะห์ performance metric ได้หลากหลาย

### 5.2 ข้อเสนอแนะ

จากการใช้งานอัลกอริทึมในงานวิจัยนี้ต้องมีการเพิ่มเติมการ label ข้อมูลโดยผู้เชี่ยวชาญ เพื่อสามารถเทียบประสิทธิภาพกับสิ่งที่อัลกอริทึมแสดงผลได้ และสามารถนำมาเปรียบเทียบกับ การใช้โมเดลการเรียนรู้ของเครื่องได้ นอกเหนือจากนี้ต้องศึกษาการสร้างคลังคำศัพท์คุณลักษณะเฉพาะเพิ่มเติม เพื่อให้เหมาะสมกับการใช้งานกับบทวิจารณ์แต่ละประเภท รวมทั้งสามารถไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพิ่มเติมคลังคำศัพท์ที่เก็บคำแสดงความรู้สึก สำหรับการจำแนกความรู้สึกที่มีต่อคุณลักษณะเพิ่มเติมได้ ทำให้สามารถวิเคราะห์ความรู้สึกเชิงบวกลบจากคุณลักษณะที่ผู้วิจัยได้ทำการทดลอง ผู้ประกอบการสามารถใช้ประโยชน์จากอัลกอริทึมเหล่านี้ได้มากยิ่งขึ้น เพื่อปรับปรุงการให้บริการ และวิเคราะห์ธุรกิจในเชิงลึก



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## เอกสารอ้างอิง

- [1] SUPRIYA V. MAHADEVKAR, BHARTI KHEMANI, SHRUTI PATIL, KETAN KOTECHEA, DEEPALI R. VORA, AJITH ABRAHAM, (Senior Member, IEEE), AND LUBNA ABDELKAREIM GABRALLA. 2022. A Review on Machine Learning Styles in Computer Vision—Techniques and Future Directions
- [2] VANNESSA DUARTE, SERGIO ZUNIGA-JARA, AND SERGIO CONTRERAS. 2022. Machine Learning and Marketing: A Systematic Literature Review
- [3] SUJATA RANI and PARTEEK KUMAR. 2021. Aspect-based Sentiment Analysis using Dependency Parsing
- [4] TARIQ ABDULLAH and AHMED AHMET. 2022. Deep Learning in Sentiment Analysis: A Survey of Recent Architectures
- [5] Vandana Yadav, Dr. Parul Verma, Prof. (Dr.) Vinodini Katiyar. 2021. E-COMMERCE PRODUCT REVIEWS USING ASPECT BASED HINDI SENTIMENT ANALYSIS
- [6] Maria Pontiki, Dimitrios Galanis, John Pavlopoulos, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis
- [7] Ashik Ahamed Aman Rafat, Mushfiqus Salehin, Fazle Rabby Khan, Syed Akhter Hossain and Sheikh Abujar. 2019. Vector Representation of Bengali Word Using Various Word Embedding Model
- [8] Wang Yue and Lei Li. 2020. Sentiment Analysis using Word2vec-CNN-BiLSTM Classification
- [9] NAILA ASLAM, FURQAN RUSTAM, ERNESTO LEE, PATRICK BERNARD WASHINGTON AND IMRAN ASHRAF. 2022.Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model
- [10] Bader M. A. Almohaimmeed. 2017. Restaurant Quality and Customer Satisfaction
- [11] Anina Troya, Reshmi Gopalakrishna Pillai, Cristian Rodriguez Rivero, Zulkuf Genc, Subhradeep Kayal, Dogu Araci. 2021. Aspect-Based Sentiment Analysis of Social Media Data With Pre-Trained Language Models
- [12] Dhanang Hadhi Sasmita, Alfian F. Wicaksono, Samuel Louvan, Mirna Adriani. 2017. Unsupervised Aspect-Based Sentiment Analysis on Indonesian Restaurant Reviews
- [13] Oanh Thi Tran, Viet The Bui. 2020. Unsupervised Aspect-Based Sentiment Analysis on Indonesian Restaurant Reviews

เอกสารนี้เป็นเอกสารที่Indonesian Restaurant Reviews เท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [14] Dhanon Leenoi. 2008. THE CONSTRUCTION OF THAI WORDNET OF 1ST ORDER ENTITY COMMON BASE CONCEPTS USING A BI-DIRECTIONAL TRANSLATION METHOD AND WITH DICTIONARIES OF DIFFERENT COMPILATIONAL APPROACHES
- [15] Chanida Kaewphet. 2020. Algorithm for Extracting Features from Product Reviews
- [16] Ravinder Ahuja, Aakarsha Chug, Shruti Kohli, Shaurya Gupta, Pratyush Ahuja. 2019. The Impact of Features Extraction on the Sentiment Analysis
- [17] Somnuk Sintuporn. 2003. Thai sentence parsing using genetic programming
- [18] Seyyed Aref Razavi. 2017. Word Embedding-based Approach to Aspect Detection for Aspect-based Summarization of Persian Customer Reviews



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ประวัติผู้เขียน

ชื่อ	นายศุภกร โกลมหทัย
วัน เดือน ปีเกิด	23 กรกฎาคม 2539
ที่อยู่ปัจจุบัน	111 ซอยอ่อนนุช 48 ถนนสุขุมวิท 77 แขวงอ่อนนุช เขตสวนหลวง กรุงเทพมหานคร 10250
ประวัติการศึกษา	(2561) วิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมสารสนเทศ เกรดเฉลี่ย 2.65 (สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง) (2566) วิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูลและการวิเคราะห์ เกรดเฉลี่ย 3.85 (สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง)



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้