

ระบบบริการให้ข้อมูลผลตอบรับของผู้บริโภค

CONSUMER FEEDBACK INFORMING SYSTEM



ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2561

ระบบบริการให้ข้อมูลผลตอบรับของผู้บริโภค

CONSUMER FEEDBACK INFORMING SYSTEM



นักสกร บุญชู
ปิติภัทร เอนกกมล

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2561

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญาานิพนธ์ปีการศึกษา 2561

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง ระบบบริการให้ข้อมูลผลตอบรับของผู้บริโภค

CONSUMER FEEDBACK INFORMING SYSTEM

ผู้จัดทำ

1. นางสาวนภัสกร บุญชู รหัสนักศึกษา 58010640

2. นายปิติภัทร เอนกกมล รหัสนักศึกษา 58010781



อาจารย์ที่ปรึกษา
(ผศ.อักรเดช วัชรระภูพงษ์)

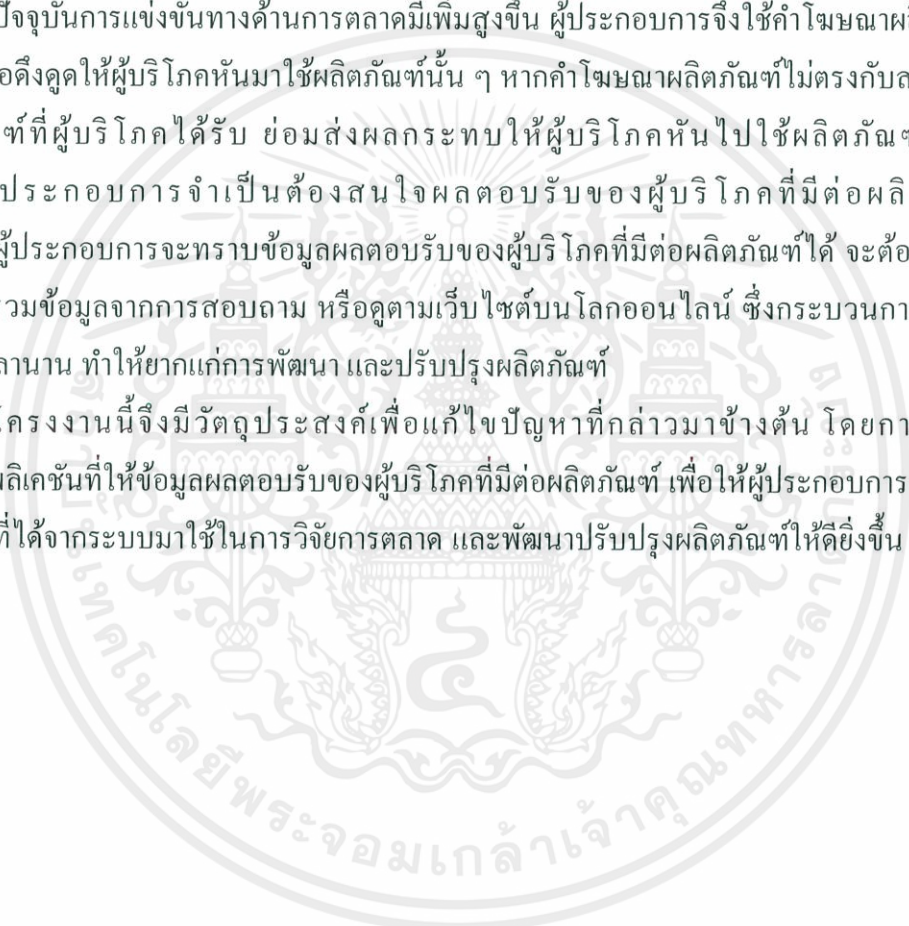
ระบบบริการให้ข้อมูลผลตอบรับของผู้บริโภค

นางสาวนภัสกร นุญชู 58010640
นายปิติภัทร เอนกกมล 58010781
ผศ.อัครเดช วัชรระภูพงษ์ อาจารย์ที่ปรึกษา
ปีการศึกษา 2561

บทคัดย่อ

ปัจจุบันการแข่งขันทางการตลาดมีเพิ่มสูงขึ้น ผู้ประกอบการจึงใช้คำโฆษณาผลิตภัณฑ์ต่าง ๆ เพื่อดึงดูดให้ผู้บริโภคหันมาใช้ผลิตภัณฑ์นั้น ๆ หากคำโฆษณาผลิตภัณฑ์ไม่ตรงกับสรรพคุณผลิตภัณฑ์ที่ผู้บริโภคได้รับ ย่อมส่งผลกระทบต่อผู้บริโภคหันไปใช้ผลิตภัณฑ์อื่นได้ ทำให้ผู้ประกอบการจำเป็นต้องสนใจผลตอบรับของผู้บริโภคที่มีต่อผลิตภัณฑ์ในการที่ผู้ประกอบการจะทราบข้อมูลผลตอบรับของผู้บริโภคที่มีต่อผลิตภัณฑ์ได้ จะต้องทำการเก็บรวบรวมข้อมูลจากการสอบถาม หรือดูตามเว็บไซต์บนโลกออนไลน์ ซึ่งกระบวนการเหล่านี้ต้องใช้เวลาานาน ทำให้ยากแก่การพัฒนา และปรับปรุงผลิตภัณฑ์

โครงการนี้จึงมีวัตถุประสงค์เพื่อแก้ไขปัญหาที่กล่าวมาข้างต้น โดยการจัดทำเว็บแอปพลิเคชันที่ให้ข้อมูลผลตอบรับของผู้บริโภคที่มีต่อผลิตภัณฑ์ เพื่อให้ผู้ประกอบการสามารถนำข้อมูลที่ได้จากระบบมาใช้ในการวิจัยการตลาด และพัฒนาปรับปรุงผลิตภัณฑ์ให้ดียิ่งขึ้น



Consumer Feedback Informing System

Ms. Napassakorn Boonchoo 58010640

Mr. Pitipat Anegkamol 58010781

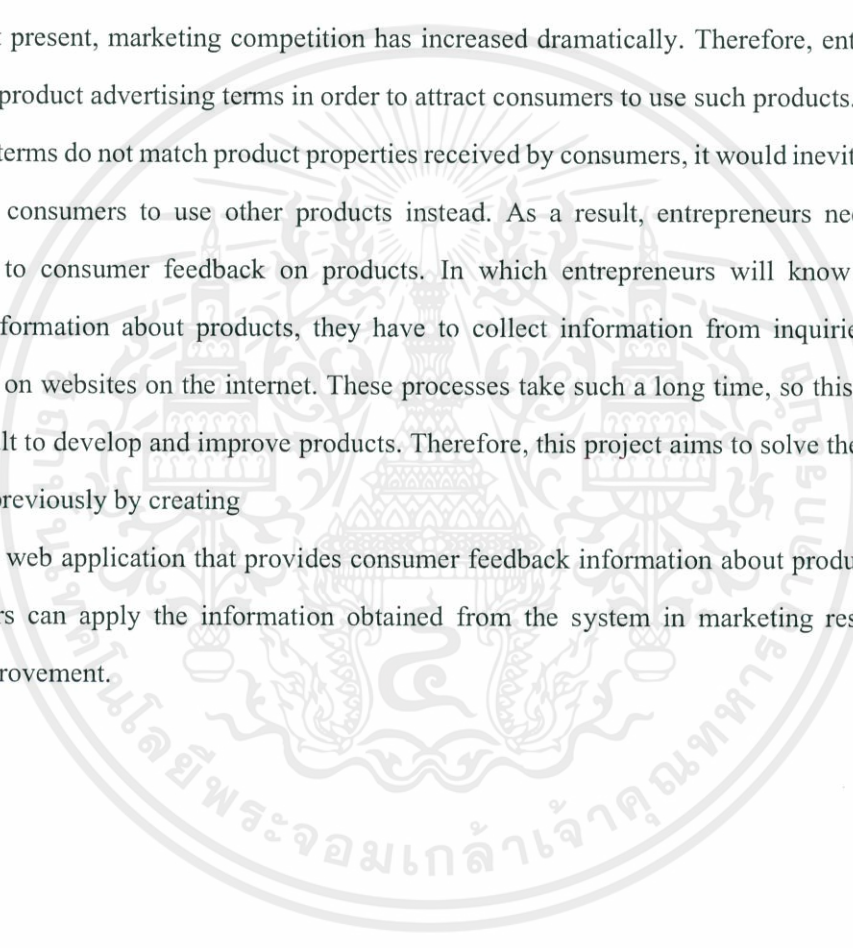
Asst.Prof. Akkradach Watcharapupong Advisor

Academic Year 2018

ABSTRACT

At present, marketing competition has increased dramatically. Therefore, entrepreneurs need to use product advertising terms in order to attract consumers to use such products. If product advertising terms do not match product properties received by consumers, it would inevitably affect decision of consumers to use other products instead. As a result, entrepreneurs need to give precedence to consumer feedback on products. In which entrepreneurs will know consumer feedback information about products, they have to collect information from inquiries or view information on websites on the internet. These processes take such a long time, so this method is quite difficult to develop and improve products. Therefore, this project aims to solve the problems mentioned previously by creating

A web application that provides consumer feedback information about products so that entrepreneurs can apply the information obtained from the system in marketing research and product improvement.



กิตติกรรมประกาศ

สำหรับปริญญาบัตรเรื่องระบบให้ข้อมูลผลตอบรับของผู้บริโภคสามารถสำเร็จลุล่วงไปได้ด้วยดี ทางคณะผู้จัดทำ ขอขอบพระคุณ ผศ. อัครเดช วัชรธัญญพงษ์ อาจารย์ที่ปรึกษาปริญญาบัตรที่คอยให้ความช่วยเหลือในเรื่องความรู้ แนวทางการพัฒนา ตลอดจนสอบถามความคืบหน้าของปริญญาบัตรอย่างสม่ำเสมอ และขอขอบพระคุณคณาจารย์ทุกท่านที่ให้ความรู้คำปรึกษาต่าง ๆ จนเป็นส่วนหนึ่งที่ทำให้ปริญญาบัตรในภาคการเรียนนี้สามารถสำเร็จลุล่วงไปได้

ขอขอบคุณเพื่อน ๆ และพี่ ๆ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังทุกคนที่คอยให้คำปรึกษากำลังใจ และความช่วยเหลือซึ่งกันและกันกับคณะผู้จัดทำมาโดยตลอด

คุณค่าและความสำเร็จของการทำปริญญาบัตรฉบับนี้ขอมอบให้กับบุคคลที่ได้กล่าวไปในข้างต้น ตลอดจนครอบครัวของคณะผู้จัดทำที่คอยแสดงความห่วงใย ถวายกำลังใจ และให้กำลังใจรวมทั้งสนับสนุนในทุก ๆ ช่วงเวลาของการดำเนินงานเสมอมา ทางคณะผู้จัดทำขอขอบพระคุณอย่างสูงไว้ ณ ที่นี้ด้วย

นภัสกร บุญชู

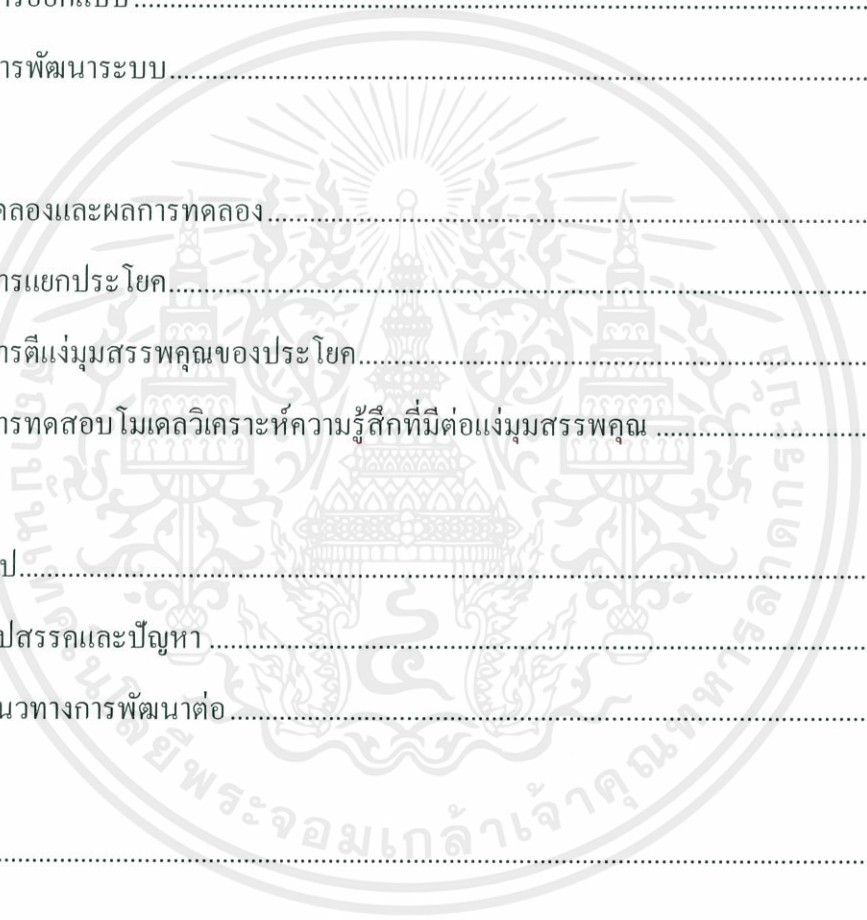
ปิติภัทร เอนกมมล

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ	IV
สารบัญตาราง	VI
สารบัญรูป.....	VII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาของปัญหา.....	1
1.2 วัตถุประสงค์ของการศึกษา.....	1
1.3 ขอบเขตของโครงการ.....	2
1.4 วิธีการดำเนินงาน.....	2
1.5 ส่วนประกอบปริญญานิพนธ์.....	4
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....	5
2.1 ภาษาไพทอน (Python).....	5
2.2 ไพไทยเอ็นแอลพี (Pythainlp).....	6
2.3 เว็บครอว์เลอร์ (Web Crawler).....	6
2.4 การเรียนรู้ของเครื่อง (Machine Learning).....	7
2.5 หลักไวยากรณ์ภาษาไทย.....	8
2.6 การประมวลผลภาษาธรรมชาติ (Natural Language Processing , NLP).....	9
2.7 เว็บแอปพลิเคชัน	13

สารบัญ (ต่อ)

	หน้า
บทที่ 3 การออกแบบและพัฒนา	18
3.1 การวิเคราะห์ความต้องการของระบบ	18
3.2 ยูสเคสไดอะแกรม (Use Case Diagram)	18
3.3 การออกแบบ	19
3.4 การพัฒนาระบบ	29
บทที่ 4 การทดลองและผลการทดลอง	36
4.1 การแยกประโยค	36
4.2 การตีแง่มุมสรรพคุณของประโยค	37
4.3 การทดสอบโมเดลวิเคราะห์ความรู้สึกที่มีต่อแง่มุมสรรพคุณ	38
บทที่ 5 บทสรุป	44
5.1 อุปสรรคและปัญหา	44
5.2 แนวทางการพัฒนาต่อ	44
บรรณานุกรม	45



สารบัญตาราง

ตาราง	หน้า
1.1 แผนการดำเนินการในภาคเรียนที่ 1	3
1.2 แผนการดำเนินการในภาคเรียนที่ 2	4
2.1 แท็กหน้าที่ของคำ	11
3.1 ตารางขอบเขตการทดลองฝึกจำแนกความรู้สึก	34
4.1 ผลลัพธ์ของขอบเขตการทดลองฝึกจำแนกความรู้สึก	39
4.2 คอนฟิวชันเมตริกซ์แสดงผลการทดสอบด้วยชุดทดสอบ (Test set) หลังจากเลือกแผนการทดลองที่มอบอัตราความแม่นยำสูงสุด	41



สารบัญรูป

รูป	หน้า
2.1 สัญลักษณ์ภาษาไพทอน.....	5
2.2 สัญลักษณ์ของไลบรารีไซคิลเดิร์น.....	5
2.3 สัญลักษณ์ของ ไพไทยเอ็นแอลพี.....	6
2.4 สัญลักษณ์ของสแครปปี.....	7
2.5 สัญลักษณ์ของบิวตี้ฟูลชูป.....	7
2.6 แนวคิดโมเดลวิวคอนโทรลเลอร์.....	15
3.1 ยูสเคสไดอะแกรม.....	18
3.2 ภาพรวมระบบ.....	19
3.3 ส่วนดึงข้อมูลบนโลกอินเทอร์เน็ต.....	19
3.4 ส่วนประมวลผลข้อมูลก่อนขึ้นฐานข้อมูล.....	20
3.5 แบบจำลองความสัมพันธ์เอนทิตีของระบบ.....	21
3.6 หน้าเข้าสู่ระบบ.....	24
3.7 หน้าลงทะเบียน.....	24
3.8 หน้าหลัก.....	25
3.9 รายละเอียดผลิตภัณฑ์ที่แสดงในหน้าหลัก.....	25
3.10 หน้ารายละเอียดผลิตภัณฑ์.....	26
3.11 แง่มุมสรรพคุณ และปริมาณการกล่าวถึงคำสำคัญต่าง ๆ บนอินเทอร์เน็ต.....	26
3.12 หน้าแสดงผลตอบรับจากผู้บริโภคในแต่ละแง่มุม.....	27
3.13 ผลตอบรับที่พึงพอใจ.....	28
3.14 หน้าประวัติการเข้าชมผลิตภัณฑ์.....	28
3.15 หน้าแสดงผลผลิตภัณฑ์ที่ชื่นชอบ.....	29
3.16 ฟังก์ชันพิจารณาแง่มุมสรรพคุณ.....	31
3.17 การฝึกจำแนกความรู้สึกต่อแง่มุมสรรพคุณ.....	32
3.18 ฟังก์ชันทำนายความรู้สึกที่มีต่อแง่มุมสรรพคุณ.....	35

บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

เนื่องจากในปัจจุบันมีผลิตภัณฑ์จำนวนมากเกิดขึ้นในท้องตลาด ส่งผลให้มีการแข่งขันกันโดยใช้คำโฆษณาต่าง ๆ เพื่อดึงดูดให้ผู้บริโภคมาใช้ผลิตภัณฑ์ ซึ่งหากคำโฆษณาผลิตภัณฑ์ไม่ตรงกับสรรพคุณของผลิตภัณฑ์ที่ผู้บริโภคได้รับ ย่อมทำให้ผู้บริโภคเปลี่ยนใจ และหันไปใช้ผลิตภัณฑ์อื่นแทน ซึ่งผู้ประกอบการสามารถทราบผลตอบรับของผู้บริโภคได้จากการสอบถาม หรือดูตามเว็บไซต์บนโลกออนไลน์เกี่ยวกับผลิตภัณฑ์ของตน แต่การที่ผู้ประกอบการนั้นจะทราบผลตอบรับของผู้บริโภคจำเป็นต้องใช้เวลาในการเก็บข้อมูลนาน ทำให้ยากแก่การพัฒนา และปรับปรุงผลิตภัณฑ์

จากปัญหาที่กล่าวมาทางคณะผู้จัดทำจึงสนใจที่จะจัดทำเว็บแอปพลิเคชันสำหรับผู้ประกอบการที่ให้ข้อมูลเกี่ยวกับผลตอบรับของผู้บริโภคที่มีต่อผลิตภัณฑ์จากโลกออนไลน์ โดยระบบจะนำผลตอบรับของผู้บริโภคที่มีต่อผลิตภัณฑ์มาจัดกลุ่มให้ตรงกับแง่มุมสรรพคุณต่าง ๆ ที่ได้จากคำโฆษณาผลิตภัณฑ์ พร้อมทั้งแบ่งกลุ่มผลตอบรับผลิตภัณฑ์เหล่านั้นว่ามีความพึงพอใจหรือไม่พึงพอใจต่อแง่มุมสรรพคุณที่กล่าวถึงอย่างไร

ระบบที่กล่าวมาข้างต้นนั้น จะช่วยให้ผู้ประกอบการได้รับข้อมูลผลตอบรับ และความคิดเห็นของผู้บริโภคที่มีต่อผลิตภัณฑ์ในรูปแบบที่เข้าใจง่าย และสามารถทราบถึงข้อดี ข้อเสียของผลิตภัณฑ์ สามารถนำข้อมูลที่ได้จากระบบมาใช้ในการวิจัยการตลาด เพื่อพัฒนาปรับปรุงผลิตภัณฑ์ และมองเห็นโอกาสทางการตลาดมากยิ่งขึ้น

1.2 วัตถุประสงค์ของการศึกษา

- 1) เพื่อนำความรู้ที่ได้ศึกษามาประยุกต์ใช้ในการพัฒนาระบบ เช่น การพัฒนาเว็บแอปพลิเคชัน (Web Application), การประมวลผลภาษาธรรมชาติ (Natural Language Processing) เป็นต้น
- 2) เพื่อให้ผู้ประกอบการทราบข้อดี และข้อเสียของผลิตภัณฑ์ เพื่อนำมาพัฒนาปรับปรุงผลิตภัณฑ์ต่อไป
- 3) เพื่อให้ผู้ประกอบการสามารถนำข้อมูลที่ได้จากเว็บไซต์มาช่วยในการวิจัยการตลาด
- 4) เพื่อให้ผู้ประกอบการเข้าใจความต้องการที่ซ่อนเร้นของผู้บริโภค และนำมาสู่ความต้องการที่แท้จริงของผู้บริโภค

- 5) เพื่อให้ผู้ประกอบการมองเห็นโอกาสทางการตลาดมากขึ้น

1.3 ขอบเขตของโครงการ

ระบบบริการให้ข้อมูลผลตอบรับผลิตภัณฑ์ของผู้บริโภค เป็นระบบที่นำแง่มุมสรรพคุณจากคำโฆษณาภาษาไทยมาจัดกลุ่มให้กับผลตอบรับผลิตภัณฑ์ที่เกี่ยวข้องกับแง่มุมสรรพคุณนั้น ๆ โดยกำหนดขอบเขตของผลิตภัณฑ์เป็นประเภทโลชั่นบำรุงผิวกาย โดยแบ่งแง่มุมออกเป็น 5 กลุ่ม ดังนี้

- 1) กลุ่มที่ 1 มีสรรพคุณทำให้ผิวขาวใส กระจ่างใส
- 2) กลุ่มที่ 2 มีสรรพคุณให้ความชุ่มชื้น ลดความหยابกร้าน
- 3) กลุ่มที่ 3 มีสรรพคุณกันแดด ปกป้องผิวจากแสงยูวี (UV)
- 4) กลุ่มที่ 4 มีสรรพคุณให้กลิ่นหอม
- 5) กลุ่มที่ 5 มีสรรพคุณเนื้อเบาบาง ซึมซาบเร็ว ไม่เหนียวเหนอะหนะ

นอกจากการจัดกลุ่มให้กับผลตอบรับผลิตภัณฑ์ที่เกี่ยวข้องกับแง่มุมสรรพคุณที่ผลิตภัณฑ์จะมอบให้กับผู้บริโภคแล้ว ระบบจะทำการจำแนกความรู้สึกของผลตอบรับผลิตภัณฑ์ให้อีกด้วย โดยแบ่งเป็นในแง่ของความพึงพอใจ (Positive) และไม่พึงพอใจ (Negative)

1.4 วิธีการดำเนินงาน

วิธีการดำเนินงานปริญญานิพนธ์นี้แบ่งออกเป็น 5 ขั้นตอนหลักดังนี้

- 1) ศึกษาปัญหา และความต้องการของผู้ใช้
- 2) ออกแบบเว็บแอปพลิเคชัน และ โมเดลให้สอดคล้องกับความต้องการของผู้ใช้
- 3) พัฒนาเว็บแอปพลิเคชัน และ โมเดลตามที่ได้ออกแบบไว้
- 4) ตรวจสอบความถูกต้องของโมเดล และการทำงานของเว็บแอปพลิเคชัน
- 5) จัดทำรายงานปริญญานิพนธ์

1.4.1 แผนการดำเนินการในภาคเรียนที่ 1

ตาราง 1.1 แผนการดำเนินการในภาคเรียนที่ 1

ลำดับที่	แผนงาน	สิงหาคม	กันยายน	ตุลาคม	พฤศจิกายน
1	ศึกษาปัญหา และความต้องการของผู้ใช้	■			
2	เลือกปัญหาที่สนใจ ซึ่งปัญหาที่สนใจคือ ระบบให้ข้อมูลผลตอบรับของผู้บริโภค		■		
3	กำหนดขอบเขต วัตถุประสงค์ของปัญหา สำรวจ และศึกษาแหล่งข้อมูลที่น่ามาใช้		■		
4	ออกแบบ Wireframe			■	
5	ออกแบบฐานข้อมูล			■	
6	ออกแบบส่วนติดต่อผู้ใช้			■	
7	เก็บข้อมูลจากแหล่งข้อมูลที่เลือก			■	
8	ออกแบบการทำงานของอัลกอริทึมเพื่อใช้ในการวิเคราะห์ความรู้สึก			■	
9	ศึกษาและพัฒนาเว็บแอปพลิเคชัน และการสร้างโมเดลในการวิเคราะห์ความรู้สึก			■	■
10	จัดทำรายงานปริญาานิพนธ์				■

ตาราง 1.2 แผนการดำเนินการในภาคเรียนที่ 2

ลำดับที่	แผนงาน	มกราคม	กุมภาพันธ์	มีนาคม	เมษายน
1	พัฒนาโมเดลวิเคราะห์ ความรู้สึก				
2	พัฒนาระบบส่วนหน้า				
3	พัฒนาระบบส่วนหลัง				
4	ปรับปรุงระบบส่วนหน้า				
5	ปรับปรุงโมเดลวิเคราะห์ ความรู้สึก				
6	นำข้อมูลผ่านการวิเคราะห์ ขึ้นไปเก็บบนฐานข้อมูล				
7	จัดทำรายงานปฏิญานិพนธ์				

1.5 ส่วนประกอบปฏิญานิพนธ์

ปฏิญานิพนธ์ฉบับนี้ได้ประกอบด้วยเนื้อหาทั้งหมด 3 บท ดังต่อไปนี้

บทที่ 1 บทนำ กล่าวถึงความสำคัญและที่มา วัตถุประสงค์ ขอบเขตการศึกษา วิธีการดำเนินการ ประโยชน์ที่คาดว่าจะได้รับ และส่วนประกอบปฏิญานิพนธ์

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง กล่าวถึงทฤษฎี และเครื่องมือที่ใช้ในการทำปฏิญานิพนธ์

บทที่ 3 การพัฒนาและการออกแบบ กล่าวถึงรายละเอียดการออกแบบ การพัฒนาโมเดล และเว็บแอปพลิเคชัน

บทที่ 4 การทดลองและผลการทดลอง การแยกประ โยค การตีแง่มุมสรรพคุณของประ โยค และ การทดสอบโมเดลวิเคราะห์ความรู้สึก

บทที่ 5 บทสรุป อุปสรรคปัญหา และแนวทางการพัฒนาต่อ

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง

2.1 ภาษาไพทอน (Python)



รูป 2.1 สัญลักษณ์ภาษาไพทอน

ภาษาไพทอน (Python) เป็นภาษาที่มีจุดเด่นในเรื่องของโอเพนซอร์ส และมีไลบรารี (Library) ที่อำนวยความสะดวกอยู่เป็นจำนวนมาก ปัจจุบันถูกนำมาพัฒนา และใช้งานอย่างแพร่หลาย ไม่ว่าจะเป็นในด้านปัญญาประดิษฐ์ (Artificial Intelligence) ด้านการเรียนรู้ของเครื่อง (Machine Learning) โดยเฉพาะอย่างยิ่งในด้านการประมวลภาษาธรรมชาติ (Natural Language Processing : NLP) ที่มีโมดูลที่รวบรวมเครื่องมือต่าง ๆ สำหรับการประมวลภาษาธรรมชาติไว้อย่างครบถ้วน

2.1.1 ไซคิตเลิร์น (Scikit-learn)



รูป 2.2 สัญลักษณ์ของไลบรารีไซคิตเลิร์น

ไซคิตเลิร์นเป็นไลบรารีสำหรับการเรียนรู้ของเครื่อง (Machine Learning) ด้วยภาษาไพทอน โดยระบบได้ใช้ในส่วนของการจำแนกข้อมูล (Classification) ซึ่งจัดเป็นการเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised Learning) สำหรับการจำแนกความรู้สึกรองข้อความ

2.2 ไฟไทยเอ็นแอลพี (Pythainlp)



รูป 2.3 สัญลักษณ์ของ ไฟไทยเอ็นแอลพี (Pythainlp)

ไฟไทยเอ็นแอลพี (Pythainlp) เป็นไลบรารีภาษาไพทอน สำหรับการประมวลภาษาธรรมชาติ โดยมุ่งเน้นไปที่การสนับสนุนภาษาไทยเป็นหลัก มีเครื่องมือสำคัญที่ถูกใช้อย่างแพร่หลาย ยกตัวอย่างเช่น ตัวตัดคำภาษาไทย (Word Tokenize) หรือตัวระบุชนิดคำภาษาไทย (Part-of-Speech) เป็นต้น

โปรแกรม 2.1 การใช้ ไฟไทยเอ็นแอลพี (Pythainlp) ในการตัดคำจากประโยคภาษาไทย

```
from pythainlp.tokenize import word_tokenize
text = "ฉันดีใจที่พวกเราได้เกิดมาบนแผ่นดินไทย"
result = word_tokenize(text)
```

2.3 เว็บครอว์เลอร์ (Web Crawler)

เว็บครอว์เลอร์ (Web Crawler) เป็นโปรแกรมสำหรับใช้เก็บรวบรวมข้อมูลบนหน้าเว็บไซต์ต่าง ๆ โดยมักจะถูกใช้ในงานประเภทเสิร์ชเอนจิน (Search Engine) เพื่อจัดทำดัชนีเว็บ (Indexing) และในงานประเภทรวบรวมข้อมูลที่ต้องการเฉพาะเจาะจง ยกตัวอย่างเช่น ทำการเก็บรวบรวมยูอาร์แอล (URL) ที่ปรากฏอยู่บนหน้าเว็บไซต์ปัจจุบันทั้งหมด หรือเก็บรวบรวมเนื้อหาบนหน้าเว็บไซต์ปัจจุบัน เป็นต้น

ในปัจจุบันภาษาไพทอนมีเฟรมเวิร์ค (Framework) หรือไลบรารีที่สามารถอำนวยความสะดวกในการสร้างเว็บครอว์เลอร์ขึ้นอย่างง่ายดาย

2.3.1 สแครปปี้ (Scrapy)



Scrapy

รูป 2.4 สัญลักษณ์ของสแครปปี้ (Scrapy)

สแครปปี้ (Scrapy) เป็นเฟรมเวิร์คสำหรับเก็บรวบรวมข้อมูลที่อยู่บนหน้าเว็บไซต์ สามารถเข้าถึง (Crawl) ไปบนเว็บไซต์ต่าง ๆ ได้อย่างรวดเร็ว และดึง โครงสร้างของหน้าเว็บไซต์นั้น ออกมาได้

2.3.2 บิวตี้ฟูลซูป (Beautifulsoup)

BeautifulSoup

รูป 2.5 สัญลักษณ์ของบิวตี้ฟูลซูป (Beautifulsoup)

บิวตี้ฟูลซูป (Beautifulsoup) เป็นไลบรารีในภาษาไพทอนสำหรับใช้ดึงข้อมูลที่อยู่บนหน้าเว็บไซต์ หรือเอชทีเอ็มแอล (HTML)

2.4 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่อง (Machine Learning : ML) เป็นกระบวนการที่ทำให้เครื่องคอมพิวเตอร์สามารถเรียนรู้ได้ด้วยตนเองปราศจากการทำงานตามลำดับคำสั่งของโปรแกรม โดยมีหลักการการทำงานคือ เมื่อมีข้อมูลเข้ามาสามารถทำนาย หรือตัดสินใจผลลัพธ์ที่ควรจะเป็นได้ โดยอ้างอิงจากการวิเคราะห์ทางสถิติ

2.4.1 การเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised learning)

การเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised learning) เป็นกระบวนการที่ทำให้เครื่องคอมพิวเตอร์สามารถเรียนรู้ได้ด้วยตนเองแบบมีมนุษย์เป็นผู้สอน เริ่มต้นจะต้องมีข้อมูลที่เตรียมพร้อมสำหรับสอนเครื่องคอมพิวเตอร์ เพื่อให้เครื่องคอมพิวเตอร์สามารถทำนายผลลัพธ์ได้ว่าควรจะตอบอย่างไร โดยมนุษย์จะต้องเป็นผู้แยกประเภทของข้อมูลเหล่านั้นก่อนเป็นอันดับแรก หรือกล่าวได้ว่า มนุษย์เป็นผู้กำหนดก่อนว่าข้อมูลที่มีรูปร่าง หรือลักษณะแบบนี้คืออะไร

หลังจากนั้นทำการแบ่งข้อมูลที่ได้แยกประเภทไว้ก่อนแล้วออกเป็นทั้งหมดสองส่วน ได้แก่ ส่วนแรกจะใช้สำหรับฝึกสอนเครื่องคอมพิวเตอร์ หรือมีชื่อเรียกว่า ข้อมูลฝึก (Training set) และส่วนถัดมาจะใช้เป็นข้อมูลไว้สำหรับทดสอบความถูกต้องที่เครื่องคอมพิวเตอร์สามารถทำนายได้ หรือมีชื่อเรียกว่า ข้อมูลทดสอบ (Test set)

2.4.2 การเรียนรู้ของเครื่องแบบไม่มีผู้สอน (Unsupervised learning)

การเรียนรู้ของเครื่องแบบไม่มีผู้สอน (Unsupervised learning) เป็นกระบวนการที่ทำให้เครื่องคอมพิวเตอร์สามารถเรียนรู้ได้ด้วยตนเองแบบไม่มีมนุษย์เป็นผู้สอน เริ่มต้นจะต้องมีข้อมูลที่เตรียมพร้อมสำหรับสอนเครื่องคอมพิวเตอร์ แต่ไม่จำเป็นจะต้องแยกประเภทของข้อมูล เพียงแค่กำหนดว่าต้องการจะแบ่งแยกข้อมูลเป็นกี่กลุ่ม (Clustering) ซึ่งแตกต่างจากการเรียนรู้ของเครื่องแบบมีผู้สอน หลังจากนั้นเครื่องคอมพิวเตอร์จะทำการจับรูปแบบของข้อมูลที่ป้อนเข้ามาว่าสิ่งใดควรอยู่ในกลุ่มเดียวกัน และแบ่งกลุ่มตามที่ได้กำหนดไว้ตั้งแต่ต้น

2.5 หลักไวยากรณ์ภาษาไทย

2.5.1 คำ

คำ คือ อักษรที่เขียน หรือพิมพ์โดยเป็นหน่วยที่เล็กที่สุดที่มีความหมายในตัว โดยชนิดของคำแบ่งออกเป็น 7 ชนิด ได้แก่

- 1) นาม เป็นคำสำหรับเรียกคน สัตว์ และสิ่งของต่าง ๆ
- 2) สรรพนาม เป็นคำที่ใช้เรียกแทนนาม เพื่อไม่ต้องกล่าวนามซ้ำอีก เช่น เธอ เขา เรา ท่าน เป็นต้น
- 3) กริยา เป็นคำที่ใช้แสดงอาการของนามหรือสรรพนาม
- 4) วิเศษณ์ เป็นคำที่ใช้บอกคุณภาพ หรือคุณลักษณะของนาม กริยา หรือวิเศษณ์ เช่น คนสวย ดิमा เก่งมาก เป็นต้น
- 5) บุพพท เป็นคำที่ใช้เชื่อมคำเข้าด้วยกัน อยู่หน้านาม สรรพนาม หรือกริยา มีคำว่า ด้วย โดย ใน เป็นต้น เช่น หนังสือในห้อง เป็นต้น

- 6) สันธาน เป็นคำที่เชื่อมประโยคให้เกี่ยวเนื่องกัน เช่น เขาชอบนก แต่ฉันชอบแมว เป็นต้น
- 7) อุทาน เป็นคำที่เปล่งออกมาเวลาตกใจ ดีใจ หรือเสียใจ เป็นต้น

2.5.2 ประโยค

ประโยค คือ ถ้อยคำที่มาเรียงกัน และมีเนื้อความบริบูรณ์ ประกอบด้วยภาคประธาน หรือผู้กระทำ และภาคแสดงคือแสดงอาการกระทำหรือกริยา โดยรูปแบบประโยคแบ่งเป็น 2 รูปแบบ ได้แก่

2.5.2.1 ประโยคไวยากรณ์

ประโยคไวยากรณ์ แบ่งประโยคเป็น 4 รูปแบบ โดยใช้โครงสร้างไวยากรณ์ คือ

- 1) ประโยคประธาน เป็นประโยคที่ประธานเป็นคำนาม หรือคำสรรพนาม และประธานทำหน้าที่เป็นผู้กระทำ วางหน้าประโยค เช่น เสือกินวัว, คนร้ายขโมยของ เป็นต้น
- 2) ประโยคกริยา เป็นประโยคที่ประธานมีคำกริยานำหน้า ซึ่งมักขึ้นต้นด้วยคำว่า มี เกิด ปรากฏ เช่น มีเสือกินวัว, ปรากฏคนร้ายขโมยของ เป็นต้น
- 3) ประโยคกรรม เป็นประโยคที่มีกรรมเป็นประธานในประโยค โดยมีกริยาลิ “ถูก” ตามหลังประธาน เช่น วัวถูกเสือกิน, ของถูกขโมยโดยคนร้าย เป็นต้น
- 4) ประโยคกริต เป็นประโยคที่มีผู้รับใช้แทรกเข้ามา เช่น นักเรียนถูกครูให้อ่านหนังสือ, ครูให้นักเรียนอ่านหนังสือ เป็นต้น

2.5.3 การเว้นวรรค

ในการเขียนภาษาไทย การเว้นช่องว่างของคำให้ถูกต้อง จะทำให้ข้อความนั้นมีความชัดเจน และเข้าใจสิ่งที่ต้องการจะสื่อได้ยิ่งขึ้น โดยการเว้นวรรคแบ่งเป็น 2 กรณี ดังต่อไปนี้

- 1) การเว้นวรรคใหญ่ เมื่อต้องการจบข้อความในแต่ละประโยค เช่น นั่งให้เรียบร้อย อย่าไขว่ห้าง
- 2) การเว้นวรรคเล็ก ใช้ในหลายกรณี ยกตัวอย่างเช่น การเว้นวรรคเมื่อเจอคำสันธาน, การเว้นวรรคหน้า และหลังเครื่องหมายต่าง ๆ , การเว้นวรรคเมื่อเจอคำอุทานวลี เป็นต้น

2.6 การประมวลผลภาษาธรรมชาติ (Natural Language Processing , NLP)

การประมวลผลภาษาทางธรรมชาติ (Natural Language Processing : NLP) เป็นเทคโนโลยีปัญญาประดิษฐ์รูปแบบหนึ่งที่พัฒนาขึ้น เพื่อให้คอมพิวเตอร์สามารถเข้าใจภาษามนุษย์ได้

2.6.1 การประมวลผลภาษาไทย

เนื่องจากภาษาไทยเป็นภาษาที่มีความยืดหยุ่น และยากที่จะทราบได้ว่าคำเหล่านั้น สิ้นสุด หรือมีขอบเขตที่ตรงไหน ดังนั้นการประมวลผลภาษาไทยจึงต้องใช้การตัดคำ และการระบุชนิดของคำ เพื่อลดความกำกวมและสับสนของคำ

2.6.1.1 การตัดคำ (Thai word segmentation)

การตัดคำ เป็นการแบ่งคำแต่ละคำในประโยคออกจากกัน เพื่อวัตถุประสงค์ต่าง ๆ เช่น การวิเคราะห์ไวยากรณ์ (Syntax analysis), การประมวลผลภาษาธรรมชาติ (Natural Language - Processing : NLP) ซึ่งการตัดคำประกอบด้วยวิธีหลัก ๆ 3 วิธี ดังต่อไปนี้

- 1) การใช้กฎ (Rule-based) เป็นวิธีที่มีความรวดเร็วในการตัดคำ และมีความถูกต้องในการตัดคำระดับพยางค์สูง แต่มีข้อจำกัดคือ ไม่สามารถตัดคำที่มีหลายพยางค์ได้ เช่น คำประสม เป็นต้น
- 2) การใช้พจนานุกรม (Dictionary) เป็นการตัดคำโดยเปรียบเทียบกับคำศัพท์ในพจนานุกรม ซึ่งวิธีนี้ใช้เวลานานกว่าการใช้กฎ แต่ให้ความถูกต้องในการตัดคำสูงกว่า
- 3) การใช้คลังข้อความ (Corpus) เป็นวิธีการตัดคำโดยอาศัยสถิติมาใช้ในการประมวลผล

ในการตัดคำจะมีการใช้เทคนิคในการตัดคำเพื่อให้คอมพิวเตอร์สามารถระบุได้ว่าขอบเขตของคำอยู่ที่ไหน โดยแบ่งออกเป็น 4 วิธีการ ดังต่อไปนี้

- 1) วิธีการเทียบคำที่ยาวที่สุด (Longest word pattern matching) เป็นการตัดคำโดยเลือกคำที่มีพยางค์มากที่สุดก่อน
- 2) วิธีการเทียบคำที่สั้นที่สุด (Shortest word pattern matching) เป็นการตัดคำโดยเลือกคำที่มีพยางค์สั้นที่สุดก่อน
- 3) วิธีการตัดคำที่ใช้ความถี่ของคำ (Word usage frequency) เป็นการตัดคำที่ใช้วิธีในข้อ 1) และ 2) ร่วมกัน โดยดูจากความถี่ของคำเหล่านั้น เพื่อแก้ปัญหาความกำกวมของคำ
- 4) วิธีการย้อนรอยกลับ (Back tracking) เป็นการตัดคำโดยเลือกคำที่มีพยางค์มากที่สุดนำไปเทียบกับคำศัพท์ในพจนานุกรม หากไม่พบคำเหล่านั้นในพจนานุกรม จะทำการย้อนไปตัดคำใหม่

2.6.2 การระบุหน้าที่ของคำ (Part-Of-Speech Tagging)

ในการตัดคำจำเป็นต้องการระบุหน้าที่ของคำ เพื่อช่วยลดความกำกวมของคำ โดยพิจารณาจากบริบทของคำ และคำที่อยู่รอบข้าง

ตาราง 2.1 แท็กหน้าที่ของคำ

ลำดับที่	POS	คำอธิบาย	ตัวอย่าง
1	NPRP	นามเฉพาะ	ประเทศไทย, นนทบุรี
2	NCNM	จำนวนนับ	1, 2, หนึ่ง, ร้อย
3	NONM	เลขแสดงลำดับ	ที่หนึ่ง, ที่สอง
4	NALN	คำวิเศษณ์ที่ทำหน้าที่เหมือนนาม	วันอาทิตย์, วันนี้, วันก่อน
5	NAJN	คำคุณศัพท์ที่ทำหน้าที่เหมือนนาม	เชิงกายภาพ, ทางสังคม
6	NLBL	นามใช้กำกับป้าย	1, 2, 3, ก, ข, ค
7	NCMN	นามทั่วไป	คอมพิวเตอร์, อาคาร, หนังสือ
8	NTTL	นามที่ใช้นำหน้าชื่อ	ยศ, ตำแหน่ง, ร้อยเอก
9	NNOM	นามวลี	ความคิดเห็น, ความดี
10	NCLT	นามที่เป็นกลุ่ม	กลุ่ม, กอง, คณะ
11	PPRS	สรรพนามที่ใช้แทนบุคคล	คุณ, ฉัน, เขา
12	PDMN	สรรพนามชี้ระยะใกล้ไกล	นี้, นั่น, นั่น
13	PDEF	สรรพนามชี้เฉพาะ	เหล่านี้, ทั้งคู่
14	PIND	สรรพนามไม่ชี้เฉพาะ	บ้าง, ทั้งหลาย
15	PNTR	สรรพนามคำถาม	ใคร, อะไร, อย่างไร
16	PREL	สรรพนามแทนนามข้างหน้า	ที่, ซึ่ง, อัน
17	VACT	กริยาที่มีการแสดงอย่างเป็นรูปธรรม	ทำงาน, ร้องเพลง, กิน
18	VSTA	กริยาที่ไม่ได้มีการแสดงอย่างเป็นรูปธรรม	เห็น, รู้, คือ
19	VATT	วิเศษณ์ที่ระบุคุณลักษณะ	เล็ก, ผอม, เย็น
20	XVBM	กริยาช่วยก่อนหน้าคำว่า “ไม่”	เกิด, เกือบ, อาจ
21	XVAM	กริยาช่วยหลังคำว่า “ไม่”	ค่อย, น่า, ได้

ตาราง 2.1 แท้กหน้าที่ของคำ (ต่อ)

ลำดับที่	POS	คำอธิบาย	ตัวอย่าง
22	XVMM	กริยาช่วยที่อยู่ได้ทั้งก่อนหน้า และหลังคำว่า “ไม่”	ควร, เคย, ต้อง
23	XVBB	กริยาช่วยขึ้นต้นประโยค	กรุณา, เชิญ, จง
24	XVAE	กริยาช่วยตามหลังกริยา	ไป, มา, ขึ้น
25	DDAN	คำบ่งชี้หลังนาม	บ้านนี้, นั้น, โนน
26	DDAC	คำบ่งชี้หลังลักษณะนาม	ดินสอแท่งนั้น, นี้, นี่
27	DDBQ	คำบ่งชี้ระหว่างนามและ ลักษณะนาม	ทั้ง, อีก, เพียง
28	DDAQ	คำบ่งชี้หลังคำที่ระบุปริมาณ	พอดี, ถ้วน
29	DIAC	คำระบุปริมาณ	อื่น, ต่าง ๆ, ไหน
30	DIBQ	คำระบุปริมาณ โดยการประมาณ	ประมาณ, ราว, เกือบ
31	DIAQ	คำระบุปริมาณ ในรูปแบบมีเศษ	กว่า, เศษ
32	DCNM	คำระบุจำนวนนับ	หนึ่งคน, เสือ 2 ตัว
33	DONM	คำระบุลำดับ	ที่หนึ่ง, ที่สุดท้าย
34	ADVN	กริยาวิเศษณ์ปกติ	เก่ง, สม่าเสมอ
35	ADVI	กริยาวิเศษณ์แสดงซ้ำ	เร็ว ๆ, ซ้ำ ๆ
36	ADVP	กริยาหลังคำอุปสรรค	โดยเร็ว
37	ADVS	กริยาวิเศษณ์ขยายประโยค	โดยปกติ, ธรรมดา
38	CNIT	ลักษณะนามระบุรูปร่าง คุณลักษณะ	ใบ, ตัว, เล่ม
39	CLTV	ลักษณะนามระบุหมวดหมู่	ฝูง, คู่, กลุ่ม
40	CMTR	ลักษณะนามในการวัด	กิโลกรัม, นาที
41	CFQC	ลักษณะนามระบุจำนวน	ครั้ง, เทียบ
42	CVBL	ลักษณะนามระบุอาการ	ม้วน, มัด
43	JCRG	สันธานเชื่อมประโยค	และ, แต่
44	JCMP	สันธานเชื่อมประโยคเชิง เปรียบเทียบ	กว่า, เหมือนกับ, เท่ากับ
45	JSBR	สันธานเชื่อมประโยคแสดง ความเป็นเหตุเป็นผล	เนื่องจาก, แม้ว่า, ถ้า

ตาราง 2.1 แท็กหน้าที่ของคำ(ต่อ)

ลำดับที่	POS	คำอธิบาย	ตัวอย่าง
46	RPRE	บุพบท	จาก, ใต้, ล่าง
47	INT	อุทาน	โอย, เออ, เอ้
48	FIXP	คำอุปสรรค	การ, ความ, อย่าง
49	EAFF	คำลงท้ายประโยคบอกเล่า	ละ, ครับ, เอะ, หน่อย
50	EITT	คำลงท้ายประโยคคำถาม	หรือ, หรือ, ไหม
51	NEG	นิเสธ	มิได้, ไม่, ไม่ได้
52	PUNC	เครื่องหมายวรรคตอน	();!

2.7 เว็บแอปพลิเคชัน

2.7.1 ภาษาจาวาสคริปต์ (JavaScript)

ภาษาจาวาสคริปต์ (JavaScript) เป็นภาษาทางคอมพิวเตอร์ที่ใช้งานบนเว็บแอปพลิเคชันต่าง ๆ เพื่อให้เว็บสามารถโต้ตอบกับผู้ใช้งานได้ดียิ่งขึ้น โดยภาษาจาวาสคริปต์เป็นภาษาสคริปต์ (Script) ซึ่งเป็นภาษาที่สามารถรันโค้ดได้โดยไม่ต้องคอมไพล์ก่อน ภาษาจาวาสคริปต์จำเป็นต้องมีตัวรันจาวาสคริปต์ (JavaScript Engine) เพื่อให้ภาษาจาวาสคริปต์สามารถรันบนเว็บแอปพลิเคชันได้

โปรแกรม 2.2 ตัวอย่างการเขียนภาษาจาวาสคริปต์

```
<html>
<head>
<title>Consumer Feedback Informing System</title>
<script language="JavaScript">
</script>
</head>
<body>Hello! </body>
</html>
```

จากโปรแกรมดังกล่าวแสดงให้เห็นว่าลักษณะการทำงานของภาษาจาวาสคริปต์เป็นภาษาสคริปต์เชิงวัตถุ ซึ่งผู้ใช้งานสามารถทำงานข้ามแพลตฟอร์มร่วมกับภาษา HTML ได้ โดยแทรกอยู่ภายในระหว่างคำสั่ง <Head> และ </Head> หรือหลังจากคำสั่ง <Body> การเขียนคำสั่งตัวพิมพ์ใหญ่และตัวพิมพ์เล็กให้ผลที่แตกต่างกัน (Case Sensitive) ซึ่งการเขียนภาษาจาวาสคริปต์เบื้องต้นมีหลักการ ดังต่อไปนี้

2.7.2 การประกาศตัวแปร

การประกาศตัวแปรในภาษาจาวาสคริปต์ประกอบด้วย var , let , const ซึ่งแต่ละการประกาศตัวแปรมีลักษณะการใช้งานที่แตกต่างกัน ดังนี้

- 1) การประกาศตัวแปรแบบ var เป็นการประกาศตัวแปรที่สามารถใช้ค่าร่วมกันได้
- 2) การประกาศตัวแปรแบบ let เป็นการประกาศตัวแปรที่สามารถใช้ค่าได้เฉพาะในส่วนที่ทำการประกาศตัวแปรไว้
- 3) การประกาศตัวแปรแบบ const เป็นการประกาศตัวแปรที่สามารถใช้ค่าได้เฉพาะในส่วนที่ทำการประกาศตัวแปรไว้ และไม่สามารถเปลี่ยนแปลงค่าได้ ยกเว้นการเปลี่ยนแปลงค่าในออบเจกต์

2.7.3 การประกาศชุดข้อมูล

การประกาศชุดของข้อมูล เป็นตัวแปรพิเศษที่สามารถเก็บค่าได้มากกว่า 1 ค่า โดยชุดข้อมูลจะมีอินเด็กซ์เริ่มต้นที่ 0 เสมอ

ตัวอย่าง 2.1 การประกาศชุดของข้อมูล

```
var name = ["A", "B", "C"];
```

2.7.4 การใช้งานฟังก์ชัน

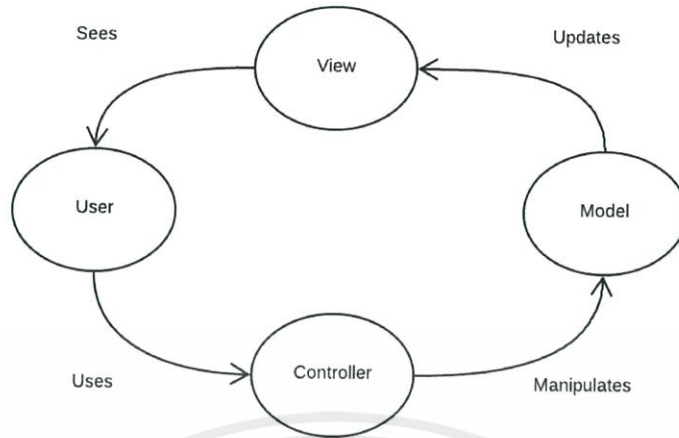
ฟังก์ชัน เป็นชุดคำสั่งที่รวมคำสั่งการทำงานไว้ และสามารถเรียกชื่อมาใช้งานตามต้องการได้

ตัวอย่าง 2.2 การใช้งานฟังก์ชัน

```
function test() {
}
test(); // เรียกใช้งาน
```

2.7.5 รีแอกต์เฟรมเวิร์ค (React Framework)

รีแอกต์เฟรมเวิร์คเป็นไลบรารีของภาษาจาวาสคริปต์ โดยใช้แนวคิดแบบโมเดลวิวคอนโทรลเลอร์ (Model View Controller , MVC) รีแอกต์เฟรมเวิร์คจะจัดการส่วนที่เป็นโมเดลหรือวิว โดยใช้คอมโพเนนต์ที่เสมือนเป็นบล็อกส่วนย่อยของเว็บ ซึ่งใช้ภาษาจาวาสคริปต์ในการระบุคุณลักษณะของคอมโพเนนต์เหล่านั้น



รูป 2.6 แนวคิดโมเดลวิวคอนโทรลเลอร์

2.7.6 เอชทีเอ็มแอล (HTML)

เอชทีเอ็มแอล (HTML) เป็นภาษาคอมพิวเตอร์รูปแบบหนึ่งที่ใช้ในการสร้างหน้าเว็บ ปัจจุบันเอชทีเอ็มแอลถูกพัฒนาเป็นรุ่นที่ 5 หรือเรียกว่า เอชทีเอ็มแอล5 (HTML5) โดยใช้โครงสร้างที่เรียกว่า “แท็ก” ในการควบคุมลักษณะการแสดงผลของข้อความ รูปภาพ และวัตถุต่าง ๆ ในแต่ละแท็กจะใช้ “แอตทริบิวต์” ซึ่งเป็นส่วนขยายสำหรับการจัดรูปแบบเพิ่มเติม

ตัวอย่าง 2.3 การเขียนภาษาเอชทีเอ็มแอล

```

<!DOCTYPE html>
<html>
<title>Consumer Feedback Informing System </title>
<body>

<h1>HTML Code</h1>
<p>This is example.</p>

</body>
</html>
  
```

2.7.7 ซีเอสเอส (CSS)

ซีเอสเอส (CSS) เป็นภาษาคอมพิวเตอร์ที่ใช้อธิบายคุณลักษณะของเอชทีเอ็มแอล (HTML) ซึ่งช่วยให้ไฟล์เอชทีเอ็มแอล (HTML) มีขนาดเล็กและเข้าใจได้ง่ายขึ้น อีกทั้งซีเอสเอส (CSS) สามารถกำหนดรูปแบบการแสดงผลจากชุดรูปแบบเดียวกันกับเอชทีเอ็มแอลหลาย ๆ ไฟล์ได้ ซึ่งทำให้สามารถปรับแก้คุณลักษณะที่ต้องการแสดงได้ง่ายยิ่งขึ้น

ตัวอย่าง 2.4 การเขียนภาษาซีเอสเอส

```
body {
  background-color: white;
}

h1 {
  font-weight: bolder;
  padding: 20px;
}

p {
  font-family: Athiti;
  font-size: 18px;
}
```

2.7.8 ไฟร์เบส (Firebase)

ไฟร์เบส (Firebase) เป็นบริการที่จัดการกับระบบเบื้องหลัง ซึ่งทำหน้าที่เป็นเอพีไอ (API) และบริการฐานข้อมูล เพื่อใช้สำหรับการพัฒนาแอปพลิเคชัน โดยบริการของไฟร์เบสสามารถใช้งานได้ฟรีและไม่เสียค่าใช้จ่าย

2.7.8.1 คลาวด์ไฟร์สโตร์ (Cloud Firestore)

คลาวด์ไฟร์สโตร์เป็นบริการสำหรับจัดเก็บ และซิงค์ข้อมูลระหว่างผู้ใช้และอุปกรณ์ โดยใช้ฐานข้อมูลโนเอสคิวแอล (NoSQL) ช่วยในการสร้างแอปพลิเคชันที่ไม่มีเซิร์ฟเวอร์

2.7.8.2 ออเทENTIเคชัน (Authentication)

ออเทENTIเคชันเป็นบริการสำหรับจัดการผู้ใช้ในวิธีที่เรียบง่าย และปลอดภัย ซึ่งมีหลายวิธีให้เลือกใช้ในการพิสูจน์ตัวตน ไม่ว่าจะเป็นการใช้อีเมลและรหัสผ่าน , การพิสูจน์ตัวตนผ่านผู้ให้บริการต่าง ๆ เช่น กูเกิ้ล (Google) , เฟสบุ๊ค (Facebook) เป็นต้น

2.7.8.3 โฮสติ้ง (Hosting)

โฮสติ้งเป็นเครื่องมือสำหรับช่วยทำเว็บโฮสติ้ง โดยอัปโหลดเนื้อหาเว็บส่งไปยังซีดีเอ็น (Content Delivery Network , CDN) ของไฟร์เบสอัตโนมัติ และมอบใบรับรองอิเล็กทรอนิกส์เอสเอสแอล (Security Socket Layer Certificates) ให้แก่เว็บแอปพลิเคชัน

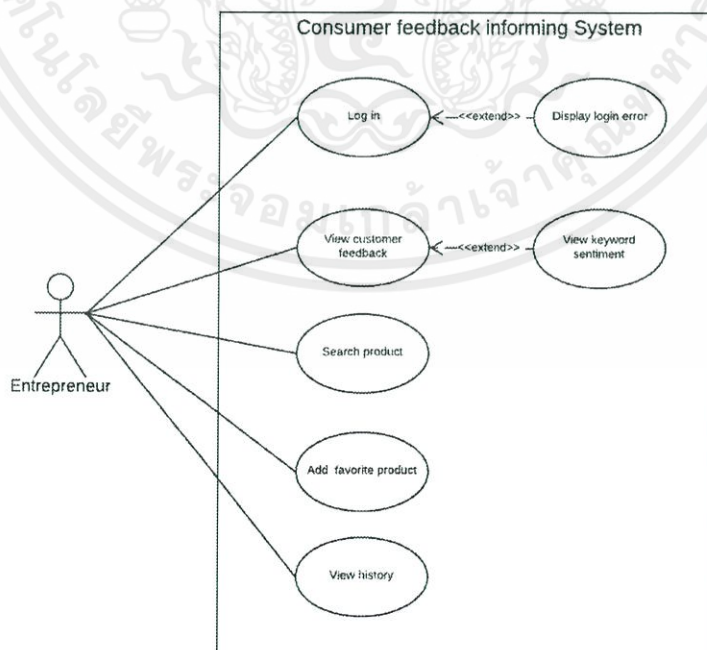
บทที่ 3

การออกแบบและพัฒนา

3.1 การวิเคราะห์ความต้องการของระบบ

- 1) ระบบสามารถจำแนกแ่งมูมสรรพคุณผลิตภัณฑ์จากคำโฆษณาของผลิตภัณฑ์ได้
- 2) ระบบสามารถวิเคราะห์ความรู้สึกของผลตอบรับจากผู้บริโภคต่อผลิตภัณฑ์ได้ 2 ประเภท ได้แก่ ความรู้สึกที่พึงพอใจ (Positive) และความรู้สึกที่ไม่พึงพอใจ (Negative)
- 3) ระบบสามารถวิเคราะห์ข้อมูลที่เป็นภาษาไทยได้
- 4) ระบบจะแบ่งแ่งมูมสรรพคุณของผลิตภัณฑ์ออกเป็น 5 กลุ่ม ได้แก่
กลุ่มที่ 1 มีสรรพคุณทำให้ผิวขาวใส กระจ่างใส
กลุ่มที่ 2 มีสรรพคุณให้ความชุ่มชื้น ลดความหยาบกร้าน
กลุ่มที่ 3 มีสรรพคุณกันแดด ปกป้องผิวจากแสงยูวี (UV)
กลุ่มที่ 4 มีสรรพคุณให้กลิ่นหอม
กลุ่มที่ 5 มีสรรพคุณเนื้อเบาบาง ซึมซาบเร็ว ไม่เหนียวเหนอะหนะ

3.2 ยูสเคสไดอะแกรม (Use Case Diagram)



รูป 3.1 ยูสเคสไดอะแกรม

3.2.1 ผู้เกี่ยวข้องกับระบบ (Actor)

ผู้ประกอบการที่ต้องการทราบผลตอบรับจากผู้บริโภคที่มีต่อผลิตภัณฑ์

3.2.2 คำอธิบายยูสเคสไดอะแกรม

- 1) Log in เข้าสู่ระบบและออกจากระบบ
- 2) Search product ค้นหาผลิตภัณฑ์
- 3) View product detail ดูรายละเอียดผลิตภัณฑ์
- 4) View history ดูประวัติการใช้งาน
- 5) Add favorite product เลือกผลิตภัณฑ์ที่สนใจ
- 6) View Keyword Sentiment ดูผลตอบรับของผู้บริโภคในแง่มุมสรรพคุณนั้น ๆ

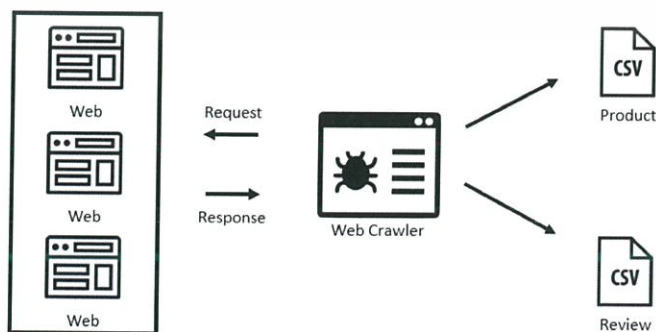
3.3 การออกแบบ

ระบบให้ข้อมูลผลตอบรับของผู้บริโภคเป็นเว็บแอปพลิเคชันซึ่ง แบ่งการทำงานออกเป็น 3 ส่วนหลักประกอบไปด้วย ส่วนดึงข้อมูลบนโลกอินเทอร์เน็ต (Web Crawler) ส่วนประมวลผลเตรียมข้อมูลก่อนขึ้นฐานข้อมูล (Processing) และส่วนแสดงผลบนเว็บแอปพลิเคชัน (Web Application)



รูป 3.2 ภาพรวมระบบ

3.3.1 ส่วนดึงข้อมูลบนโลกอินเทอร์เน็ต (Web Crawler)



รูป 3.3 ส่วนดึงข้อมูลบนโลกอินเทอร์เน็ต

สำหรับในส่วนนี้จะเป็นส่วนที่จะใช้สำหรับดึงข้อมูลบนโลกอินเทอร์เน็ต (Internet) เริ่มต้นจะมีการกำหนดปลายทางของรายชื่อเว็บไซต์ที่จะดึงข้อมูล ซึ่งเป็นเว็บไซต์ชื่อดังที่มีผู้ใช้งานเข้ามาแสดงผลตอบรับผลิตภัณฑ์เป็นจำนวนมากโดยในส่วนนี้มีเป้าหมายคือ ทำการดึงข้อมูลที่จำเป็นสำหรับระบบทั้งหมดออกมา ซึ่งได้แก่

1) ผลิตภัณฑ์ (Product) ประกอบไปด้วย

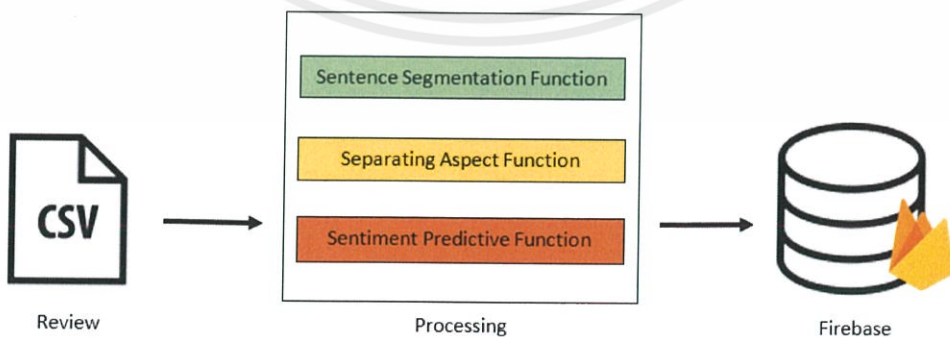
- 1.1) หมายเลขไอดีของผลิตภัณฑ์ (Product id)
- 1.2) ชื่อแบรนด์สินค้าของผลิตภัณฑ์ (Brand name)
- 1.3) ชื่อสินค้าของผลิตภัณฑ์ (Item name)
- 1.4) คำอธิบายสรรพคุณของผลิตภัณฑ์ (Description)
- 1.5) แหล่ง URL ที่มาของผลิตภัณฑ์ (URL)

2) ผลตอบรับผลิตภัณฑ์ (Review) ประกอบไปด้วย

- 2.1) หมายเลขไอดีของผลตอบรับผลิตภัณฑ์ (Review id)
- 2.2) ชื่อแบรนด์สินค้าที่แสดงผลตอบรับผลิตภัณฑ์ (Brand name)
- 2.3) ชื่อสินค้าที่แสดงผลตอบรับผลิตภัณฑ์ (Item name)
- 2.4) ชื่อผู้เขียนผลตอบรับผลิตภัณฑ์ (Author)
- 2.5) ข้อความผลตอบรับผลิตภัณฑ์ (Review)
- 2.6) วันที่เขียนผลตอบรับผลิตภัณฑ์ (Date)
- 2.7) แหล่ง URL ที่มาของผลตอบรับผลิตภัณฑ์ (URL)

โดยข้อมูลที่มีความสำคัญมากที่สุด เพื่อนำไปแยกแ่งมุมมองสรรพคุณ และวิเคราะห์ความรู้สึกต่อแ่งมุมมองนั้น ๆ คือในส่วนของคำอธิบายสรรพคุณของผลิตภัณฑ์ (Description) และข้อความผลตอบรับผลิตภัณฑ์ (Review)

3.3.2 ส่วนประมวลผลข้อมูลก่อนขึ้นฐานข้อมูล (Processing)

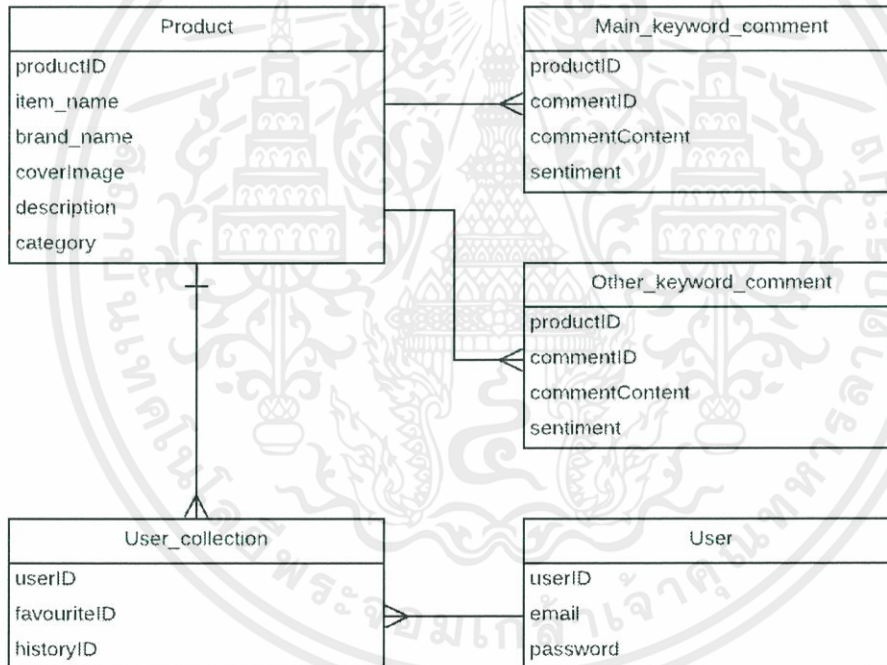


รูป 3.4 ส่วนประมวลผลข้อมูลก่อนขึ้นฐานข้อมูล

โดยเริ่มต้นทำการค้นหาผลตอบรับผลิตภัณฑ์ที่ตรงกับผลิตภัณฑ์ทั้งหมดที่ระบบมีอยู่ เมื่อได้ผลตอบรับผลิตภัณฑ์ที่มีผลิตภัณฑ์ตรงกับที่ระบบมีอยู่แล้ว ก็จะนำผลตอบรับของผลิตภัณฑ์เหล่านั้นเข้าสู่ 3 ส่วนของการประมวลผลข้อมูลก่อนขึ้นฐานข้อมูล อันได้แก่ ส่วนแยกประโยค (Sentence Segmentation) ส่วนพิจารณาแง่มุมสรรพคุณ (Aspects of Product Properties) และส่วนวิเคราะห์ความรู้สึกต่อแง่มุมสรรพคุณ (Sentiment Predictive) และสุดท้ายจะทำการจัดเก็บเป็นโครงสร้างของคิกชันนารีในภาษาไพทอนที่เหมาะสม พร้อมนำขึ้นไปเก็บบนฐานข้อมูลต่อไป

3.3.2.1 ฐานข้อมูล (Database)

ฐานข้อมูลในระบบนี้จะต้องเก็บข้อมูลของผลิตภัณฑ์ และผลตอบรับของผลิตภัณฑ์ทั้งหมดเอาไว้ เพื่อนำไปใช้ในส่วนของการแยกแง่มุมสรรพคุณ และวิเคราะห์ความรู้สึกต่อแง่มุมสรรพคุณนั้น ๆ



รูป 3.5 แบบจำลองความสัมพันธ์เอนทิตีของระบบ

3.3.2.2 ส่วนแยกประโยค (Sentence Segmentation)

3.3.2.2.1 ฟังก์ชันแยกประโยค (Sentence Segmentation function)

ฟังก์ชันแยกประโยค (Sentence Segmentation function) เป็นส่วนที่แยกประโยคผลตอบรับผลิตภัณฑ์ให้อยู่ในรูปของประโยคย่อย โดยอ้างอิงจากไวยากรณ์ภาษาไทย เพื่อให้ได้ประโยคที่มีแง่มุมสรรพคุณอยู่เพียงแง่มุมเดียว ซึ่งการแยกประโยคจะถูกใช้กับประโยค

ผลตอบรับผลิตภัณฑ์ก่อนนำไปแยกแ่งมุ่ม และเข้าสู่โมเดลวิเคราะห์ความรู้สึกต่อแ่งมุ่มสรรพคุณต่อไป

3.3.2.3 ส่วนพิจารณาแ่งมุ่มสรรพคุณ (Separate aspect of product properties)

ระบบได้ถูกออกแบบให้มีลักษณะการทำงานดังรูป 3.6 ซึ่งมีรายละเอียดดังต่อไปนี้

3.3.2.3.1 กลุ่มคำของแ่งมุ่มสรรพคุณทั้งหมด (Aspects of Product Properties)

กลุ่มคำของแ่งมุ่มสรรพคุณทั้งหมด (Aspects of Product Properties) เป็นส่วนที่รวบรวมคำที่เกี่ยวกับแ่งมุ่มสรรพคุณในกลุ่มของผลิตภัณฑ์โลชั่นบำรุงผิวขาวไว้แล้ว ซึ่งแบ่งได้ออกเป็นทั้งหมด 5 กลุ่มหลัก ได้แก่

1) กลุ่มที่ 1 กลุ่มสรรพคุณให้ผิวขาวใส

ตัวอย่างคำที่เกี่ยวข้องได้แก่ ขาว, กระจ่างใส, ใส

2) กลุ่มที่ 2 กลุ่มสรรพคุณให้ความชุ่มชื้น ลดความหยابกร้าน

ตัวอย่างคำที่เกี่ยวข้องได้แก่ ชุ่มชื้น, ชุ่มชื่น, หยาบกร้าน, นุ่ม

3) กลุ่มที่ 3 กลุ่มสรรพคุณให้กันแดด

ตัวอย่างคำที่เกี่ยวข้องได้แก่ ยูวี, กันแดด, แสงแดด

4) กลุ่มที่ 4 กลุ่มสรรพคุณให้กลิ่นหอม

ตัวอย่างคำที่เกี่ยวข้องได้แก่ กลิ่น, หอม, ชวน

5) กลุ่มที่ 5 กลุ่มสรรพคุณให้เนื้อบางเบา ไม่เหนียวเหนอะหนะ

ตัวอย่างคำที่เกี่ยวข้องได้แก่ บางเบา, ซึมซาบ, เหนียว

3.3.2.3.2 ฟังก์ชันพิจารณาแ่งมุ่มสรรพคุณ (Separating Aspect of Product

Properties Function)

ฟังก์ชันพิจารณาแ่งมุ่มสรรพคุณ (Separating Aspect of Product Properties Function) เป็นส่วนที่สร้างฟังก์ชันที่มีหน้าที่พิจารณาว่าคำอธิบายสรรพคุณ หรือประโยคผลตอบรับผลิตภัณฑ์เป็นแ่งมุ่มสรรพคุณใด

- 1) สำหรับประโยคคำอธิบายสรรพคุณ จะทำการตัดคำ และพิจารณาจากคำทั้งหมดที่ปรากฏอยู่ในประโยคคำอธิบายสรรพคุณว่าอยู่ในกลุ่มคำของแ่งมุ่มสรรพคุณกลุ่มใด โดยจะมีข้อมูลนำเข้าไปเป็นประโยคคำอธิบายสรรพคุณ และมีผลลัพธ์เป็นรายการของแ่งมุ่มของสรรพคุณที่กล่าวถึงทั้งหมดในประโยคนั้น

- 2) สำหรับประโยคผลตอบรับผลิตภัณฑ์ จะทำการตัดคำพิจารณาจากคำทั้งหมดที่ปรากฏอยู่ในประโยคผลตอบรับผลิตภัณฑ์ว่าอยู่ในกลุ่มคำของแ่งมมสรรพคุณกลุ่มใด โดยจะมีข้อมูลนำเข้าเป็นประโยคย่อยที่ผ่านการตัดประโยคจากประโยคของผลตอบรับผลิตภัณฑ์มาเรียบร้อยแล้ว และมีผลลัพธ์เป็นแ่งมมสรรพคุณที่กล่าวถึงในประโยคนั้น

3.3.2.4 ส่วนวิเคราะห์ความรู้สึกต่อแ่งมมสรรพคุณ (Sentiment Analytic)

3.3.2.4.1 การฝึกจำแนกความรู้สึกต่อแ่งมมสรรพคุณ (Sentiment Trainer)

ส่วนที่จะใช้ทำการสร้างการเรียนรู้ให้กับเครื่องคอมพิวเตอร์ในเรื่องของความรู้สึกที่มีต่อแ่งมมสรรพคุณนั้น โดยจะต้องนำประโยคย่อยที่ผ่านการตัดประโยคจากประโยคทั้งหมดของผลตอบรับผลิตภัณฑ์ ซึ่งมีแ่งมมสรรพคุณเป็นของตนเองมาตัดคำ และเก็บอยู่ในรูปของถุงคำ (Bag of words) โดยทำการคัดเลือกคำที่มีผลต่อการวิเคราะห์ความรู้สึกเป็นคำสำคัญ หลังจากนั้นก็จะนำข้อมูลทั้งหมดที่ผ่านการคัดกรอง และจัดประเภทความรู้สึกมาแล้วมาเป็นข้อมูลเพื่อใช้ฝึกการเรียนรู้ของเครื่องคอมพิวเตอร์ และสร้างเป็นโมเดลสำหรับเรียกใช้งานต่อไป

3.3.2.4.2 ฟังก์ชันทำนายความรู้สึกต่อแ่งมมสรรพคุณนั้น (Sentiment Predictive

Function)

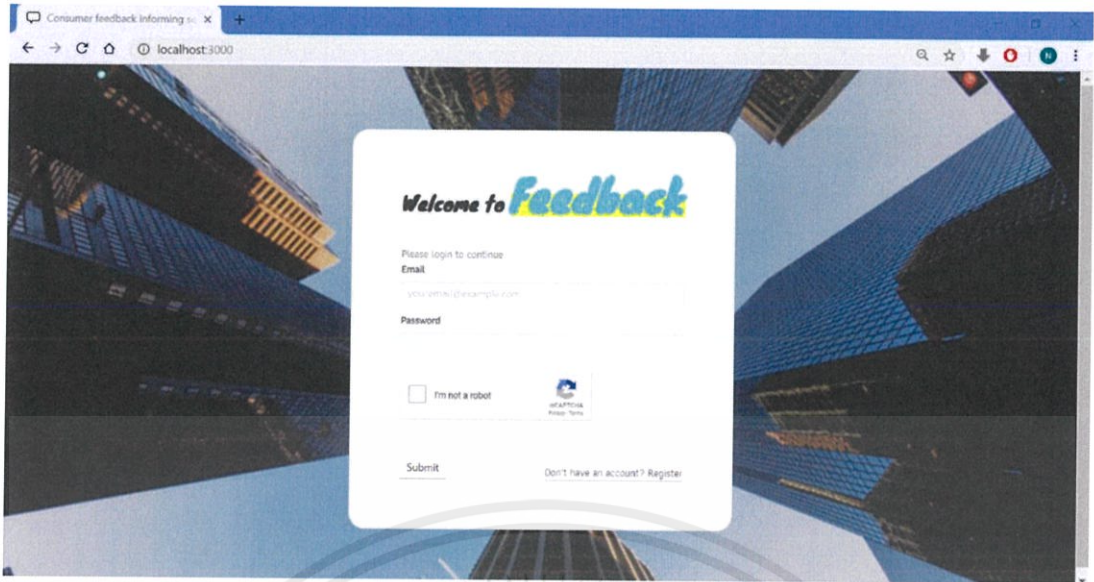
ฟังก์ชันทำนายความรู้สึกต่อแ่งมมสรรพคุณนั้น (Sentiment Predictive Function) เป็นส่วนที่นำโมเดลที่ผ่านการฝึกจำแนกความรู้สึกมาใช้ โดยในส่วนนี้จะต้องทำการสร้างฟังก์ชันที่สามารถใส่ข้อมูลเป็นประโยคย่อยที่ผ่านการตัดประโยคจากประโยคของผลตอบรับผลิตภัณฑ์ และมีแ่งมมสรรพคุณที่กล่าวถึงในประโยค โดยผลลัพธ์จะออกมาเป็นความรู้สึกของประโยคย่อยนั้น

3.3.3 ส่วนแสดงผลบนเว็บแอปพลิเคชัน (Web Application)

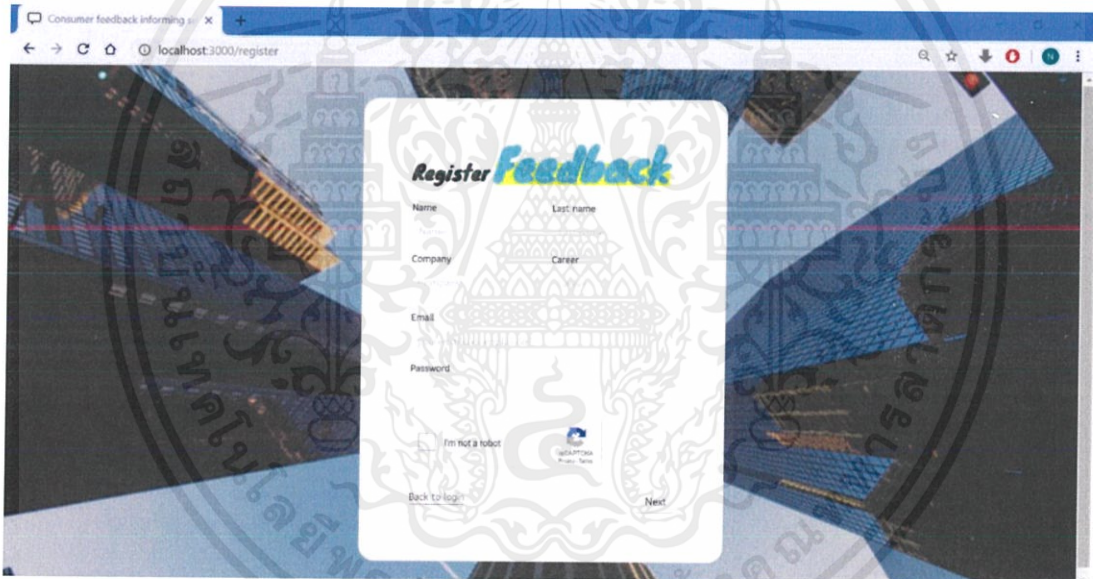
3.3.3.1 ส่วนติดต่อผู้ใช้งาน (User Interface)

ระบบให้ข้อมูลผลตอบรับจากผู้บริโภค มีส่วนติดต่อผู้ใช้งาน ดังนี้

- 1) หน้าเข้าสู่ระบบ ผู้ใช้สามารถเข้าชมเว็บแอปพลิเคชันได้โดยใช้อีเมลและรหัสผ่านกรอกในหน้าเข้าสู่ระบบเพื่อยืนยันตัวตน
- 2) หน้าลงทะเบียน สำหรับผู้ใช้ใหม่จะต้องกรอกข้อมูลเพื่อทำการลงทะเบียน เมื่อผู้ใช้กรอกข้อมูลเสร็จสิ้นแล้ว ระบบจะทำการส่งอีเมลเพื่อให้ผู้ใช้ยืนยันตัวตน เมื่อผู้ใช้ยืนยันตัวตนผ่านอีเมลแล้ว จะสามารถเข้าใช้งานเว็บแอปพลิเคชันได้โดยการเข้าสู่ระบบที่หน้าเข้าสู่ระบบ



รูป 3.6 หน้าเข้าสู่ระบบ



รูป 3.7 หน้าลงทะเบียน

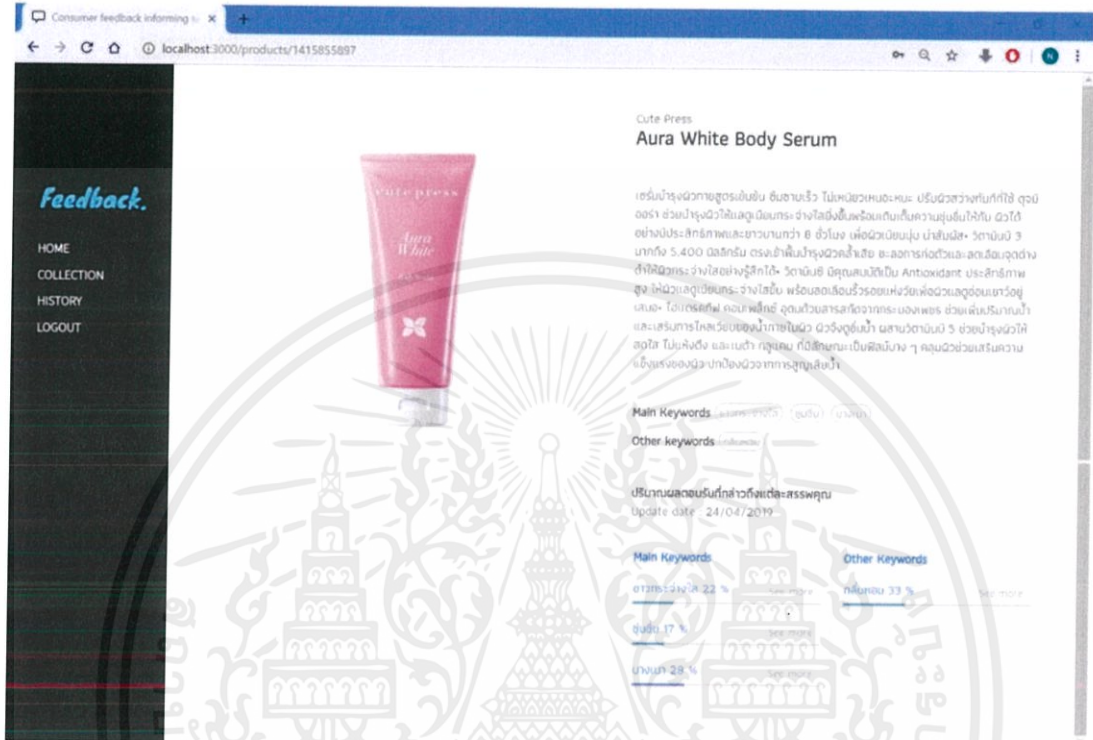
- 3) หน้าหลัก หน้าหลัก ผู้ใช้สามารถค้นหา และดูรายละเอียดผลิตภัณฑ์แต่ละชิ้นได้ นอกจากนี้ผู้ใช้ยังสามารถเลือกผลิตภัณฑ์ที่ชื่นชอบโดยการกดปุ่มเก็บ (Collect) และสามารถดูผลิตภัณฑ์ที่เก็บได้ที่หน้าคอลเล็กชัน (Collection)



รูป 3.8 หน้าหลัก

รูป 3.9 รายละเอียดผลิตภัณฑ์ที่แสดงในหน้าหลัก

- 4) หน้ารายละเอียดผลิตภัณฑ์ แสดงรายละเอียดของผลิตภัณฑ์ แก้มุมสรรพคุณที่วิเคราะห์ได้จากคำโฆษณาผลิตภัณฑ์ และปริมาณการกล่าวถึงคำสำคัญต่างๆ บนอินเทอร์เน็ต

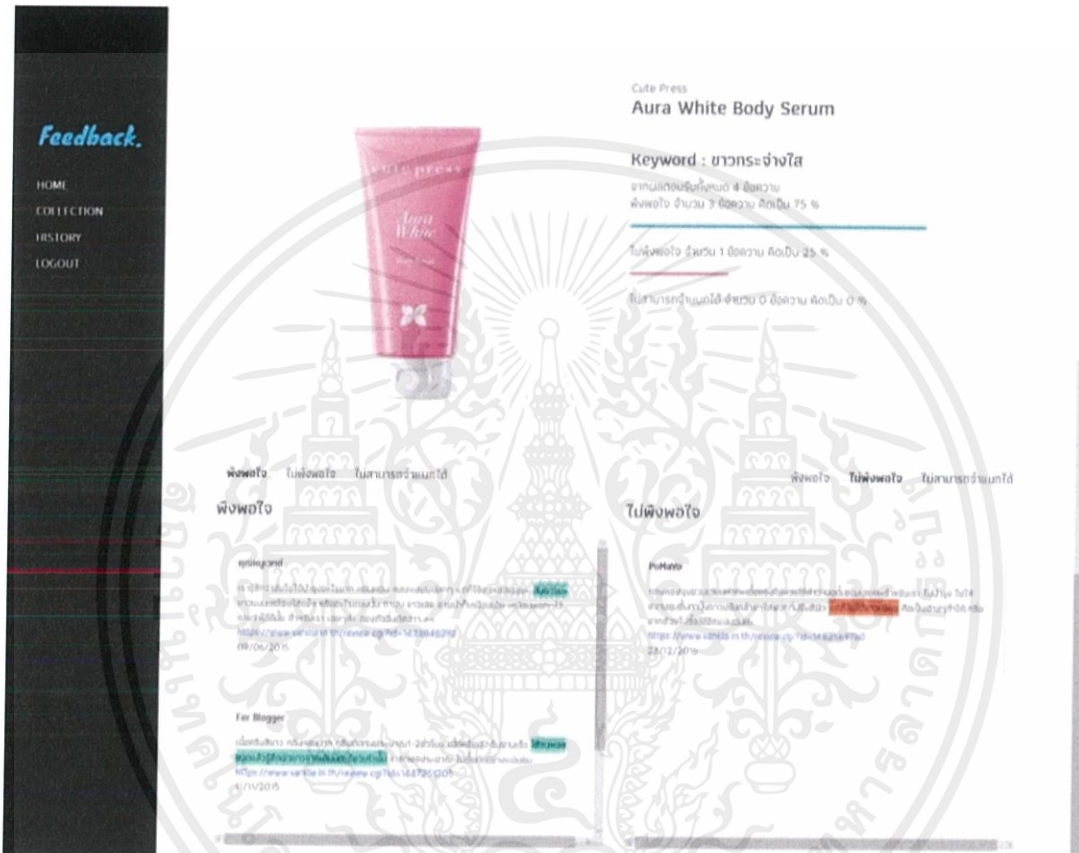


รูป 3.10 หน้ารายละเอียดผลิตภัณฑ์

Main product properties		ขาวกระจ่างใส	ชุ่มชื้น	บำรุงผิว
Other product properties		กลิ่นหอม		
ปริมาณผลตอบรับที่กล่าวถึงแต่ละสรรพคุณ				
Update date : 24/04/2019				
Main product properties		Other product properties		
ขาวกระจ่างใส 18 %	See more	กลิ่นหอม 35 %	See more	
ชุ่มชื้น 18 %	See more			
บำรุงผิว 29 %	See more			

รูป 3.11 แก้มุมสรรพคุณ และปริมาณการกล่าวถึงคำสำคัญต่างๆ บนอินเทอร์เน็ต

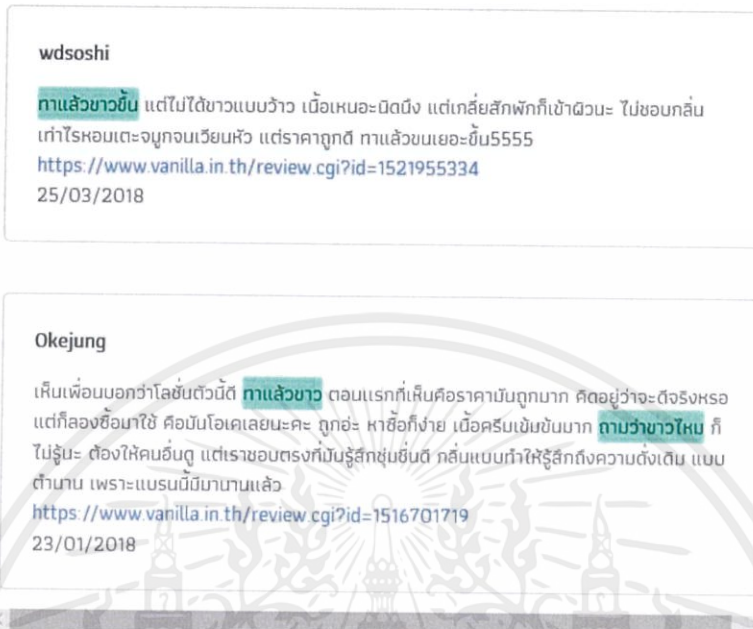
- 5) หน้าแสดงผลตอบรับ แสดงผลตอบรับที่พึงพอใจ ไม่พึงพอใจ และไม่สามารถจำแนกได้ ซึ่งจะมีการเน้นข้อความประโยคที่วิเคราะห์ความรู้สึก โดยใช้สีเขียวแสดงผลตอบรับที่พึงพอใจ สีแดงสำหรับผลตอบรับที่ไม่พึงพอใจ และสีเหลืองสำหรับผลตอบรับที่ไม่สามารถจำแนกได้ โดยผู้ใช้สามารถเลือกเปรียบเทียบผลตอบรับที่พึงพอใจ ไม่พึงพอใจ และไม่สามารถจำแนกได้



รูป 3.12 หน้าแสดงผลตอบรับจากผู้บริโภคในแต่ละแง่มุม

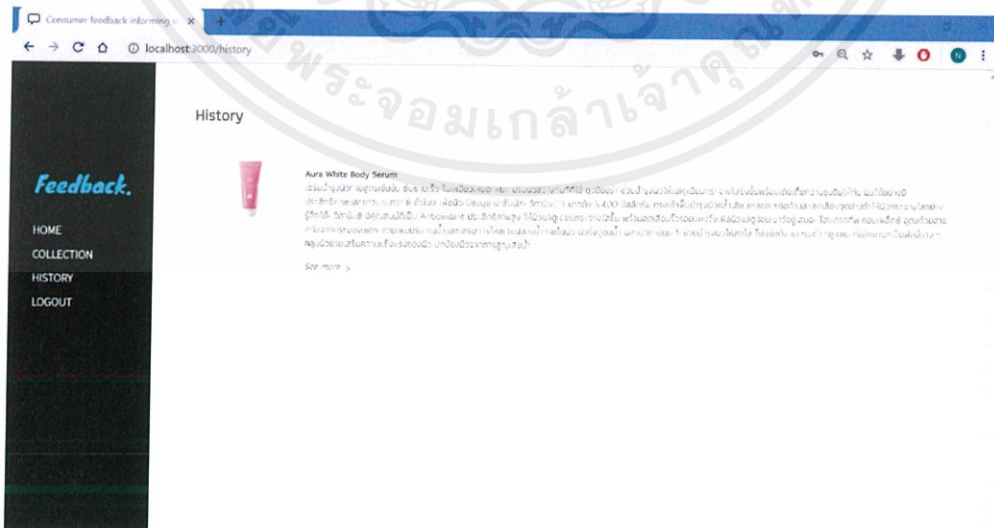
พึงพอใจ ไม่พึงพอใจ ไม่สามารถจำแนกได้

พึงพอใจ



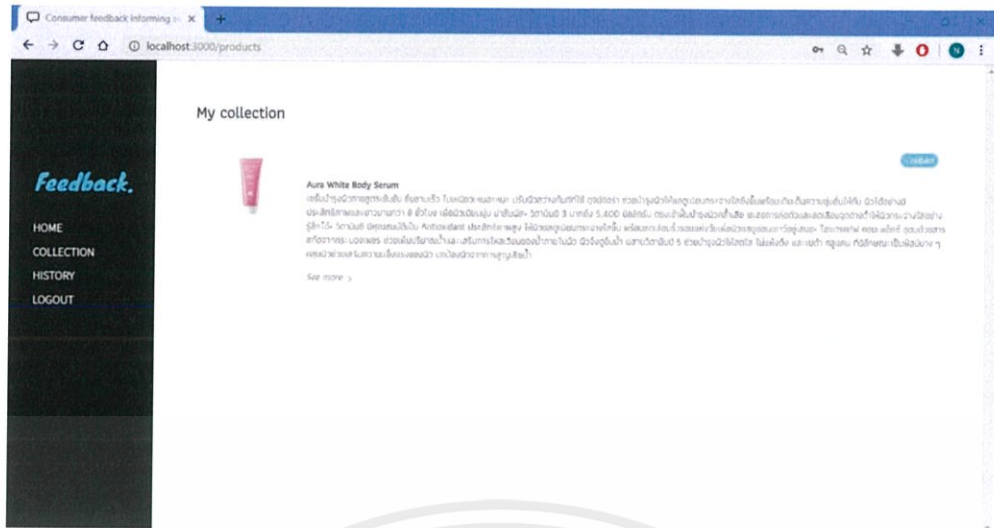
รูป 3.13 ผลตอบรับที่พึงพอใจ

- 6) หน้าประวัติการเข้าชม แสดงผลิตภัณฑ์ที่ผู้ใช้ได้เคยเข้าชม
- 7) หน้าแสดงผลิตภัณฑ์ที่ชื่นชอบ แสดงผลิตภัณฑ์ที่ผู้ใช้เพิ่มในผลิตภัณฑ์ที่ชื่นชอบไว้ สามารถลบผลิตภัณฑ์ที่ชื่นชอบออกได้



รูป 3.14 หน้าประวัติการเข้าชมผลิตภัณฑ์

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 3.15 หน้าแสดงผลผลิตภัณฑ์ที่ชื่นชอบ

3.4 การพัฒนาระบบ

3.4.1 การกำหนดคำตอบของความรู้สึกของผลตอบรับผลิตภัณฑ์

ในการทำระบบให้ข้อมูลผลตอบรับจากผู้บริโภค จะนำประโยชน์จากผลตอบรับผลิตภัณฑ์ของผู้บริโภคจำนวนหนึ่ง มาแยกประโยชน์ให้เป็นประโยชน์ย่อย และให้อาสาสมัครเพศชายกับหญิงอายุตั้งแต่ 18 ปีขึ้นไปจำนวน 5 คน ที่มีความคุ้นเคยกับภาษาไทยโดยให้อาสาสมัครกลุ่มนี้มาทำแบบสอบถามเพื่อตอบคำถามว่า ประโยคใดบ้างที่สื่อถึงความรู้สึกพึงพอใจ หรือความรู้สึกไม่พึงพอใจ คำตอบที่อาสาสมัครกำหนดให้อย่างน้อยจาก 3 ใน 5 จะถือเป็นเอกลักษณ์ แล้วกำหนดให้เป็นคำตอบความรู้สึกของประโยคที่แยกออกมาแล้วนั้นทันที

3.4.2 การดึงข้อมูลจากเว็บไซต์

ส่วนที่จะใช้สำหรับดึงข้อมูลบน โลกอินเทอร์เน็ต และข้อมูลผลตอบรับของผลิตภัณฑ์จะเป็นข้อมูลที่มาจากเว็บไซต์ที่มีความน่าเชื่อถือ โดยมีเกณฑ์พิจารณาความน่าเชื่อถือจากทางเว็บไซต์จะต้องมีการป้องกันการสแปมผลตอบรับ มีการยืนยันตัวตนว่าเป็นสมาชิกของเว็บไซต์ ไม่ใช่บุคคลทั่วไปที่เข้ามาแสดงผลตอบรับผลิตภัณฑ์ และมีการกำหนดจำนวนตัวอักษรที่สามารถป้อนได้ของผลตอบรับผลิตภัณฑ์ เป็นต้น

ในการดึงข้อมูลจากเว็บไซต์จะใช้เฟรมเวิร์คสแครปปี้ (Scrapy) สำหรับรีเควสต์ไปยังยูอาร์แอล (URL) ต่าง ๆ ที่กำหนดไว้ตั้ง และใช้บิวตี้ฟูลซุป (Beautifulsoup) เพื่อที่จะดึงข้อมูลที่ถูกรับอยู่ในรูปเอชทีเอ็มแอล (HTML) ที่สนใจออกมา เช่น ข้อมูลผลิตภัณฑ์ หรือข้อมูลผลตอบรับผลิตภัณฑ์ รวมทั้งยูอาร์แอลที่ต้องการจะรีเควสต์ต่อไปด้วย

โปรแกรม 3.1 การรีควีสต์เว็บไซต์ตามที่กำหนดไว้ และตรวจเช็คว่ามีข้อมูลอยู่บนหน้าเว็บไซต์นั้น

```
class TestSpider(scrapy.Spider):
    name = 'Test'
    url = 'https://www.cosmenet.in.th'

    def start_requests(self):
        yield scrapy.Request(
            '{} /ranking/1947/Body-Cream'.format(self.url),
            self.parse,
        )
    def parse(self, response):
        soup = BeautifulSoup(response.body, 'lxml')
        if soup.find(class_='content main-content'):
            print("Pass")
        else:
            print("Failed")
```

ข้อมูลเหล่านั้นที่ถูกดึงมาจะถูกเก็บไว้ในรูปแบบของไฟล์ซีเอสวี (CSV) และถูกอ่านเข้ามาเป็นตัวแปรดิคชันนารีในภาษาไพทอน เพื่อให้ง่ายต่อวิเคราะห์ต่อไป

ตัวอย่าง 3.1 การเก็บข้อมูลผลิตภัณฑ์ในรูปแบบของตัวแปรดิคชันนารี

```
OrderedDict([
    ('Review_ID', 'vani_143826663'),
    ('Date', '30/07/2015'),
    ('Brand_name', 'Vaseline'),
    ('Item_name', 'Healthy White UV Lightening Vitamin B3 Triple Sunscreen'),
    ('Review', 'ใช้แล้วรู้สึกผิวชุ่มชื้น ซึมเร็ว กลิ่นค่อนข้างหอม'),
    ('Author', 'PNCJY'),
    ('URL', 'https://www.vanilla.in.th/review.cgi?id=1438266633')])
```

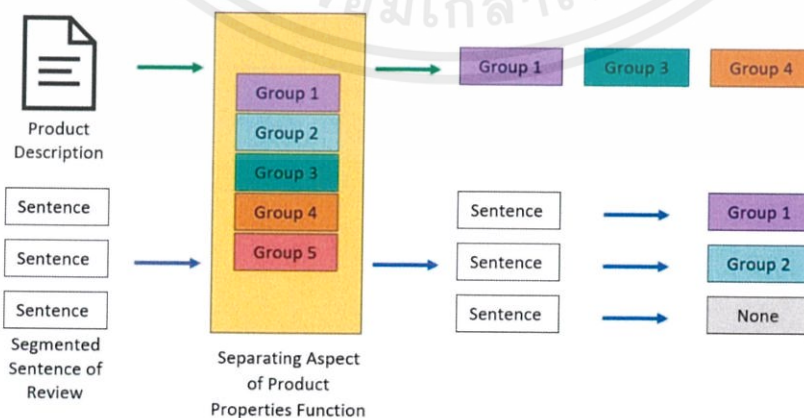
ตัวอย่าง 3.2 การเก็บข้อมูลผลตอบรับผลิตภัณฑ์ในรูปแบบของตัวแปรดิคชันนารี

```
OrderedDict ([
  ('PID', '1347965713'),
  ('Brand_name', 'Vaseline'),
  ('Item_name', 'Healthy White UV Lightening'),
  ('Img_link',
'http://www.vaseline.in/Resources/Images/Vaseline-
200ml_400x470%20(2)_tcm2887-1149803.png'),
  ('Description', 'วาสลีน เฮอร์รี่ ไวท์ สกิน โล้ท์เทนนิ่ง โลชั่น โลชั่นบำรุงผิว
ที่ทำให้ผิวคุณกระจ่างใส ดูสมำเสมอทั่วเรือนร่าง ที่มีพร้อมทั้งวิตามิน B3 ช่วยให้ผิว
กระจ่างใสอย่างเป็นธรรมชาติ มีโยเกิร์ต โปรตีน ช่วยบำรุงให้ผิวชุ่มชื้น และทริเปปไทด์
ชั้นสกรีน ช่วยปกป้องผิวจากอันตรายจากรังสียูวี'),
  ('URL',
'https://www.vanilla.in.th//detail.cgi?pid=1347965
713')])
```

3.4.3 ฟังก์ชันแยกประโยค (Sentence Segmentation Function)

ฟังก์ชันแยกประโยค (Sentence Segmentation Function) จะนำประโยคมาแยกออกเป็นประโยคย่อย โดยจะทำการตัดคำ (Word Tokenize) ในแต่ละประโยค และระบุหน้าที่ของคำ (Part of Speech Tagging) เหล่านั้น จากนั้นนำชุดคำที่ได้รับการระบุหน้าที่ของแต่ละประโยคมาตรวจสอบเงื่อนไขทางไวยากรณ์ภาษาไทยเพื่อทำการแยกประโยค เช่น เมื่อเจอคำสันธานจะทำการแยกประโยคเป็นประโยคย่อย เป็นต้น

3.4.4 ฟังก์ชันพิจารณาแง่มุมสรรพคุณ (Separating Aspect of Product Properties Function)

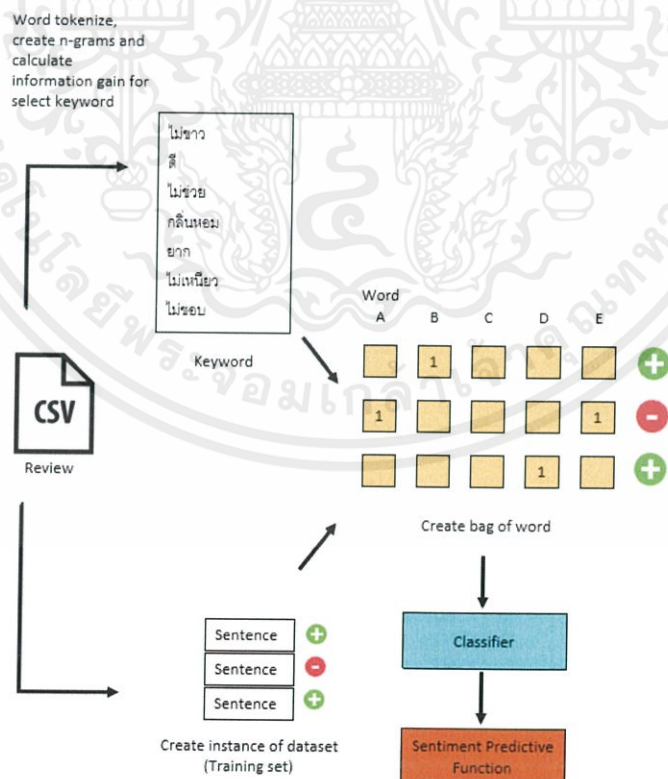


รูป 3.16 ฟังก์ชันพิจารณาแง่มุมสรรพคุณ

สำหรับฟังก์ชันพิจารณาแง่มุมสรรพคุณ (Separating Aspect of Product Properties Function) นั้นจะทำการดึงค่าจากไฟล์ที่สร้างไว้ที่เก็บชุดค่าของแง่มุมสรรพคุณ เพื่อนำมาพิจารณา สำหรับข้อมูลขาเข้าที่จะเข้ามาในฟังก์ชัน โดยข้อมูลขาเข้าสามารถเป็นได้ทั้งในส่วนของประโยค คำอธิบายสรรพคุณ และประโยคผลตอบรับผลิตภัณฑ์

- 1) สำหรับประโยคคำอธิบายสรรพคุณจะทำการตัดคำด้วย ไพไทยเอ็นแอลพี และนำเอาเรื่องของเอ็นแกรม เข้ามาช่วยในการตรวจสอบคำ จากนั้นจะทำการพิจารณาจากคำทั้งหมดที่ปรากฏอยู่ในประโยคคำอธิบายสรรพคุณว่าอยู่ในกลุ่มสรรพคุณใดบ้าง มีผลลัพธ์เป็นรายการของแง่มุมของสรรพคุณที่กล่าวถึงทั้งหมดในประโยคนั้น
- 2) สำหรับประโยคผลตอบรับผลิตภัณฑ์จะทำการตัดคำด้วย ไพไทยเอ็นแอลพี และนำเอ็นแกรม เข้ามาช่วยเช่นเดียวกัน แต่ข้อมูลนำเข้าของประโยคผลตอบรับผลิตภัณฑ์จะแตกต่างจากประโยคคำอธิบายสรรพคุณ โดยข้อมูลนำเข้าจะต้องเป็นประโยคย่อยที่ผ่านการตัดประโยคจากประโยคของผลตอบรับผลิตภัณฑ์มาเรียบร้อยแล้ว และมีผลลัพธ์เป็นแง่มุมสรรพคุณของประโยคนั้น

3.4.5 การฝึกจำแนกความรู้สึกต่อแง่มุมสรรพคุณ (Sentiment Trainer)



รูป 3.17 การฝึกจำแนกความรู้สึกต่อแง่มุมสรรพคุณ

สำหรับการฝึกจำแนกความรู้สึกต่อแง่มุมสรรพคุณที่มีต่อผลิตภัณฑ์ (Sentiment Trainer) ซึ่งจะแบ่งออกเป็นความรู้สึกพึงพอใจ และความรู้สึกไม่พึงพอใจ เริ่มจากนำกลุ่มประโยคของผลตอบรับผลิตภัณฑ์มาผ่านฟังก์ชันแยกประโยคแล้วมาตัดคำจากทุกประโยคย่อยด้วยไลบรารี Python เอนแอลพี จากนั้นทำการกรองคำที่ไม่มี ความหมายต่อการจำแนกความรู้สึกทิ้งไป (Stop word) และจะหาคำสำคัญ (Feature) ของประโยคที่แยกเหล่านั้น โดยหากมีคำสำคัญก็จะถูกเก็บอยู่ในรูปของถุงคำ (Bag of words) และเตรียมนำเข้าสู่ตัวจำแนก (Classifier) โดยส่วนนี้จะใช้ตัวจำแนกเป็น Logistic Regression และ Bernoulli Naïve Bays

3.4.5.1 การตรวจสอบคำสำคัญด้วยเอ็นแกรม (N-grams detection)

การใช้ไลบรารีตัดคำภาษาไทยบางครั้งจะตัดคำสำคัญออกจากกัน เพราะคำสำคัญนั้นเกิดจากการรวมกันของคำมากกว่า 1 คำ ดังนั้นจะไม่สามารถตรวจสอบคำสำคัญที่ต้องการในกรณีนี้ได้ การแก้ปัญหาจากกรณีนี้จะนำวิธีการของการจับเอ็นแกรม (N-grams) เพื่อรวมคำไว้เป็นกลุ่ม โดยระบบนี้จะทำการรวมคำตั้งแต่แบบยูนิแกรม (unigram) ไปจนถึงสูงสุดถึงแบบไตรแกรม (trigram)

ยกตัวอย่างเช่น ถ้าต้องการคำว่า “ไม่ขาวกระจ่างใส” การใช้ไลบรารีตัดคำของ Python เอนแอลพีจะได้แบบยูนิแกรม (unigram) เพียงอย่างเดียวเท่านั้น และจะมีลักษณะของคำเป็น “ใช้” - “แล้ว” - “ไม่” - “ขาว” - “กระจ่างใส” และถ้าทำการรวมคำแบบไตรแกรม (trigram) มาช่วยจะได้ลักษณะของคำเป็น “ใช้แล้วไม่” - “แล้วไม่ขาว” - “ไม่ขาวกระจ่างใส” จะทำให้สามารถตรวจสอบคำว่า “ไม่ขาวกระจ่างใส” ได้

3.4.5.2 การเลือกคำสำคัญโดยใช้ค่าเกนความรู้ (Information gain)

นอกจากการตรวจสอบคำสำคัญด้วยเอ็นแกรมแล้ว จะมีวิธีการเลือกคำสำคัญ (Keyword) โดยการคำนวณหาค่าเกนความรู้ (Information gain) เพื่อตัดสินใจว่าคำสำคัญใดหากนำมาใช้จำแนกความรู้สึกแล้วจะมีประสิทธิภาพที่ดี โดยนำประโยคที่ให้อาสาสมัครกำหนดคำตอบความรู้สึก มาตัดคำด้วยไลบรารี Python เอนแอลพี และทำการคำนวณค่าเกนความรู้เพื่อใช้เป็นตัวตัดสินใจว่าคำใดควรนำมาใช้เป็นคำสำคัญของระบบ และนำไปใช้ในเลือกคำใส่ถุงคำต่อไป

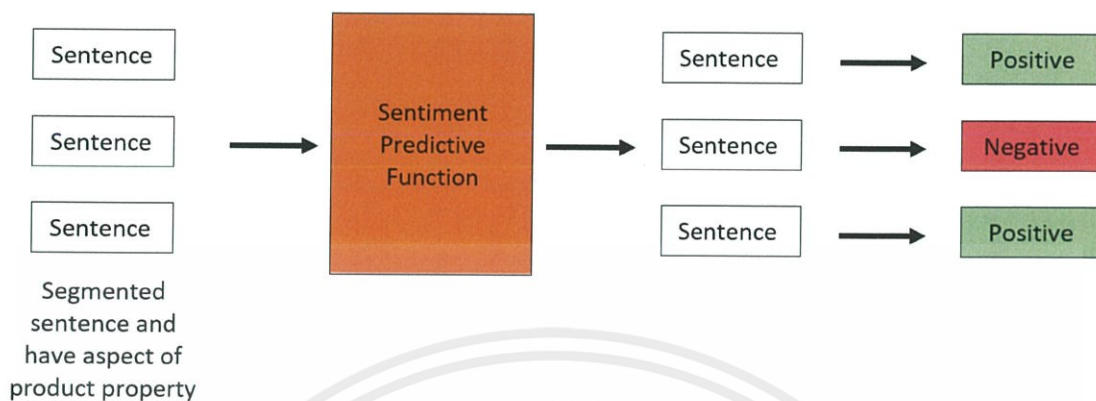
3.4.5.3 ขอบเขตการทดลองฝึกจำแนกความรู้สึก (Experiment landscape)

สำหรับขอบเขตของการทดลองจะเป็นไปตามตารางที่ 3.1 ซึ่งผลการทดลองที่มอบอัตราความแม่นยำสูงที่สุด ก็ถูกเลือกในการนำมาใช้เป็นโมเดลสำหรับจำแนกความรู้สึกต่อแง่มุมสรรพคุณ

ตารางที่ 3.1 ตารางขอบเขตการทดลองฝึกจำแนกความรู้สึก

Experiment Set	Feature set	Experiment Set	Feature set
1. Logistic Regression U	Unigrams	9. Logistic Regression UT	Uni + Trigrams
2. Bernoulli Naive Bayes U	Unigrams	10. Bernoulli Naive Bayes UT	Uni + Trigrams
3. Logistic Regression B	Bigrams	11. Logistic Regression BT	Bi + Trigrams
4. Bernoulli Naive Bayes B	Bigrams	12. Bernoulli Naive Bayes BT	Bi + Trigrams
5. Logistic Regression T	Trigrams	13. Logistic Regression UBT	Uni + Bigrams + Trigrams
6. Bernoulli Naive Bayes T	Trigrams	14. Bernoulli Naive Bayes UBT	Uni + Bigrams + Trigrams
7. Logistic Regression UB	Uni + Bigrams		
8. Bernoulli Naive Bayes UB	Uni + Bigrams		

3.4.6 ฟังก์ชันทำนายความรู้สึกที่มีต่อแง่มุมสรรพคุณ (Sentiment Predictive Function)



รูป 3.18 ฟังก์ชันทำนายความรู้สึกที่มีต่อแง่มุมสรรพคุณ

ในส่วนของฟังก์ชันทำนายความรู้สึกที่มีต่อแง่มุมสรรพคุณ จะรับข้อมูลเข้าเป็นประโยคที่ผ่านการแยกประโยคจากประโยคผลตอบรับผลิตภัณฑ์พร้อมทั้งมีแง่มุมสรรพคุณที่สื่อถึงมาแล้ว โดยผลลัพธ์จะถูกแสดงเป็นความรู้สึกที่มีต่อแง่มุมสรรพคุณนั้นว่าพอใจหรือไม่พอใจ

ตัวอย่าง 3.7 การเรียกใช้ฟังก์ชันทำนายความรู้สึกที่มีต่อแง่มุมสรรพคุณ

Example : `predict("กลิ่นไม่ฉุนหอมกำลังดี")`

ตัวอย่าง 3.8 ผลลัพธ์จากเรียกใช้ฟังก์ชันทำนายความรู้สึกที่มีต่อแง่มุมสรรพคุณ

Result : `pos`

บทที่ 4

การทดลองและผลการทดลอง

4.1 การแยกประโยค

ส่วนการแยกประโยคจะนำข้อความผลตอบรับที่เก็บรวบรวมมาจากเว็บไซต์รีวิวผลิตภัณฑ์ต่างๆ มาแยกประโยคเป็นประโยคย่อยๆ การแยกประโยคจะอ้างอิงหลักเกณฑ์การเว้นวรรคจากสำนักงานราชบัณฑิตยสภา รวมทั้งพิจารณาชนิดของคำแต่ละคำในประโยค เช่น คำที่เป็นคำสันธาน เป็นต้น หากข้อความใดมีเงื่อนไขตรงตามที่กำหนด จะทำการแยกประโยค หากไม่ตรงตามเงื่อนไขที่กำหนดจะนำไปต่อกับข้อความก่อนหน้า และทำการตรวจสอบข้อความต่อไป

ตัวอย่าง 4.1 ประโยคที่ต้องการแยกประโยค

หอมมาก เหมือนขนมเลย กลิ่นเหมือนเพียวริคูวดสีแดงชาขาวกลิ่นมิชเบอร์รี่ผสมกับชูกัส กลิ่นราสเบอร์รี่ ทาก่อนนอนแล้วหลับสบายเลยคะ มิวแต่ดมกลิ่นเลยไม่ได้สังเกตเรื่องอื่นเท่าไร เออคร่าวๆคือ เนื้อโลชั่นชั้นและค่อนข้างเหนียว ซึมเร็วดี ทาแล้วไม่เหนอะหนะ ชุ่มชื้นพอประมาณ ไม่ถึงกับนุ่มรู้สึกได้ ราคาไม่แพง ซื้อมาทาเล่นๆหอมๆ

ตัวอย่าง 4.2 ผลลัพธ์การแยกประโยคจากประโยคตัวอย่าง 4.1

- หอมมาก
- เหมือนขนมเลย
- กลิ่นเหมือนเพียวริคูวดสีแดงชาขาวกลิ่นมิชเบอร์รี่ผสมกับชูกัสกลิ่นราสเบอร์รี่
- ทาก่อนนอนแล้วหลับสบายเลยคะ
- มิว
- แต่ดมกลิ่นเลยไม่ได้สังเกตเรื่องอื่นเท่าไร
- เออคร่าวๆคือ
- เนื้อโลชั่นชั้น
- และค่อนข้างเหนียว
- ซึมเร็วดี
- ทาแล้วไม่เหนอะหนะ
- ชุ่มชื้นพอประมาณ

ตัวอย่าง 4.2 ผลลัพธ์การแยกประโยชน์จากประโยชน์ตัวอย่าง 4.1 (ต่อ)

- ไม่ถึงกับไม่รู้สึกรสได้
- ราคาไม่แพง
- ซื้อมาทานเล่นๆหอมๆ

การแยกประโยชน์จะพิจารณาจากการเว้นวรรคและคำสันธานเป็นหลัก ประโยคที่ไม่มีการใช้คำสันธาน ไม่มีการเว้นวรรค หรือเว้นวรรคไม่ถูกต้องตามหลักเกณฑ์ อาจทำให้ประสิทธิภาพในการแยกประโยคลดลงได้

4.2 การตีแ่งมมสรรพคุณของประโยค

สำหรับการพิจารณาแ่งมมสรรพคุณของประโยค จะทำการพิจารณาทั้งในประโยคคำอธิบายสรรพคุณ และผลตอบรับผลิตภัณฑ์ ซึ่งแบ่งได้ออกเป็นทั้งหมด 5 กลุ่มหลัก ได้แก่

กลุ่มที่ 1 กลุ่มสรรพคุณให้ผิวขาวใส

ตัวอย่างคำที่เกี่ยวข้องได้แก่ ขาว, กระจ่างใส, ใส

กลุ่มที่ 2 กลุ่มสรรพคุณให้ความชุ่มชื้น ลดความหยาบกร้าน

ตัวอย่างคำที่เกี่ยวข้องได้แก่ ชุ่มชื้น, ชุ่มชื้น

กลุ่มที่ 3 กลุ่มสรรพคุณให้กันแดด

ตัวอย่างคำที่เกี่ยวข้องได้แก่ ยูวี, กันแดด

กลุ่มที่ 4 กลุ่มสรรพคุณให้กลิ่นหอม

ตัวอย่างคำที่เกี่ยวข้องได้แก่ กลิ่น, หอม, ฉุน

กลุ่มที่ 5 กลุ่มสรรพคุณให้เนื้อบางเบา ไม่เหนียวเหนอะหนะ

ตัวอย่างคำที่เกี่ยวข้องได้แก่ บางเบา, ซึมซาบ, เหนียว

1. สำหรับประโยคคำอธิบายสรรพคุณ จะมีข้อมูลนำเข้าเป็นประโยคคำอธิบายสรรพคุณ และมีผลลัพธ์เป็นชุดของแ่งมมของสรรพคุณที่มีทั้งหมดในประโยคนั้น ๆ

ตัวอย่าง 4.3 ประโยคคำอธิบายสรรพคุณที่ต้องการพิจารณาแ่งมมสรรพคุณ

วาสลิน เฮลตี้ ไวท์ สกิน โลไลท์เทนนิ่ง โลชั่น โลชั่นบำรุงผิวที่ทำให้ผิวคุณกระจ่างใส ดูสม่่าเสมอทั่วเรือนร่าง ที่มีพร้อมทั้งวิตามิน B3 ช่วยให้ผิวกระจ่างใสอย่างเป็นธรรมชาติ มีโยเกิร์ต โปรตีน ช่วยบำรุงให้ผิวชุ่มชื้น และทริเปปไทด์ ชันสกรีน ช่วยปกป้องผิวจากอันตรายจากรังสียูวี

ตัวอย่าง 4.4 ผลลัพธ์การพิจารณาแง่มุมสรรพคุณจากประโยคตัวอย่าง 4.3

- ขาวกระจ่างใส
- ชุ่มชื้น
- กันแดด

2. สำหรับประโยคผลตอบรับผลิตภัณฑ์ จะมีข้อมูลนำเข้าเป็นประโยคย่อยที่ผ่านการแยกประโยคจากประโยคของผลตอบรับผลิตภัณฑ์มาเรียบร้อยแล้ว และผลลัพธ์เป็นแง่มุมสรรพคุณเพียงแง่มุมเดียวเท่านั้น

ตัวอย่าง 4.5 ประโยคผลตอบรับผลิตภัณฑ์ที่ต้องการพิจารณาแง่มุมสรรพคุณ

กลิ่น ไม่นุ่มหอมกำลังดี

ตัวอย่าง 4.6 ผลลัพธ์การพิจารณาแง่มุมสรรพคุณจากประโยคตัวอย่าง 4.5

กลิ่นหอม

สำหรับทั้งประโยคคำอธิบายสรรพคุณ และประโยคผลตอบรับผลิตภัณฑ์ฟังก์ชันสามารถทำงานได้เป็นอย่างดี เพราะมีการกำหนดขอบเขตของกลุ่มสรรพคุณที่ต้องการ และที่เกี่ยวข้องกับคำสำคัญไว้หมดแล้ว

ในกรณีของประโยคผลตอบรับผลิตภัณฑ์ หากพิจารณาแล้วไม่ได้สื่อถึงแง่มุมสรรพคุณของกลุ่มใดออกมา ก็จะไม่ถูกนำไปเข้าสู่ฟังก์ชันทำนายความรู้สึก

4.3 การทดสอบโมเดลวิเคราะห์ความรู้สึกที่มีต่อแง่มุมสรรพคุณ

หัวข้อนี้จะพูดถึงการแบ่งข้อมูล และทดสอบความแม่นยำของโมเดลจำแนกความรู้สึก ที่มีต่อแง่มุมสรรพคุณ สำหรับการแบ่งข้อมูลจะมีการแบ่งออกเป็นชุดข้อมูลที่ใช้ฝึก (Trained set) และชุดข้อมูลที่ใช้ทดสอบ (Test set)

4.3.1 การแบ่งข้อมูล

จากการให้อาสาสมัครเพศชายและหญิงอายุตั้งแต่ 18 ปีขึ้นไปจำนวน 5 คน ที่มีความคุ้นเคยกับภาษาไทยโดยให้อาสาสมัครกลุ่มนี้มาทำแบบสอบถามเพื่อตอบคำถามว่า ประโยคที่ทำการเลือกขึ้นมา แต่ละประโยคนั้นแสดงถึงความรู้สึกในแง่บวก หรือความรู้สึกในแง่ลบ ซึ่งได้ปริมาณประโยคทั้งหมดจำนวน 2200 ประโยค โดยแบ่งเป็น 1760 ประโยคเป็นข้อมูลที่ใช้ฝึก (training set) และ 440 ประโยคเป็นข้อมูลที่ใช้ทดสอบ (test set) และนำข้อมูลที่ใช้ฝึกไปสร้างโมเดลสำหรับจำแนกความรู้สึกขึ้นมา

4.3.2 ผลลัพธ์ของขอบเขตการทดลองฝึกจำแนกความรู้สึก

ตามขอบเขตการทดลองในตารางที่ 3.1 ได้ผลลัพธ์การทดลองเป็นไปตามตารางที่ 4.1 ต่อไปนี้ จากการทดลองในข้อที่ 1. จนถึงข้อ 14. ข้อแรก จะสังเกตเห็นได้ว่าในข้อที่ 7. และ 13. จะมออบอัตราความแม่นยำสูงที่สุดเป็นอันดับหนึ่ง และอันดับสองตามลำดับ โดยมีจุดที่แตกต่างกันคือ ในข้อ 7. เป็นการรวมกันของยูนิแกรม (unigrams) กับไบแกรม (bigrams) ส่วนในข้อที่ 13. เป็นการรวมกันของยูนิแกรม, ไบแกรม และไตรแกรม (trigrams) จึงได้นำทั้ง 2 การทดลองมาฝึกจำแนกความรู้สึกอีกครั้ง โดยนำค่าเกณฑ์ความรู้มาช่วยพิจารณาในการเลือกคำสำคัญ

หลังจากนำกำหนดค่าเกณฑ์ความรู้มาเพื่อสำหรับตัดสินใจ ถึงแม้ว่าในการทดลองที่ 15. จะมออบค่าความแม่นยำที่สูงกว่าข้อที่ 16. แต่ระบบได้มีการเลือกการทดลองข้อ 16. ที่มออบอัตราความแม่นยำอยู่ที่ 94% เป็นโมเดลสำหรับจำแนกความรู้สึก

เนื่องจากอัตราความแม่นยำ (test accuracy) ต่างกันประมาณ 0.01 และจำนวนการตรวจจับคำของข้อ 16. ที่มากกว่า การจับคำในการทดลองข้อที่ 16. นั้น สามารถจับได้ตั้งแต่ยูนิแกรม จนถึงไตรแกรม ทำให้สามารถแก้ไขปัญหาในกรณีพบเจอคำที่เกิดการรวมตัวของ 3 คำได้

ตาราง 4.1 ผลลัพธ์ของขอบเขตการทดลองฝึกจำแนกความรู้สึก

Experiment Set	Feature set	Validation Accuracy (5-fold)	Test Accuracy
1. Logistic Regression U	Unigrams	0.7436	0.7295
2. Bernoulli Naive Bayes U	Unigrams	0.7309	0.7
3. Logistic Regression B	Bigrams	0.9155	0.9022

ตาราง 4.1 ผลลัพธ์ของขอบเขตการทดลองฝึกจำแนกความรู้สึก (ต่อ)

Experiment Set	Feature set	Validation Accuracy (5-fold)	Test Accuracy
4. Bernoulli Naive Bayes B	Bigrams	0.9154	0.8818
5. Logistic Regression T	Trigrams	0.6986	0.6795
6. Bernoulli Naive Bayes T	Trigrams	0.6977	0.675
7. Logistic Regression UB	Uni + Bigrams	0.9491	0.9455
8. Bernoulli Naive Bayes UB	Uni + Bigrams	0.8627	0.8523
9. Logistic Regression UT	Uni + Trigrams	0.805	0.8022
10. Bernoulli Naive Bayes UT	Uni + Trigrams	0.7886	0.7727
11. Logistic Regression BT	Bi + Trigrams	0.9127	0.9023
12. Bernoulli Naive Bayes BT	Bi + Trigrams	0.8995	0.8795
13. Logistic Regression UBT	Uni + Bigrams + Trigrams	0.9481	0.9386
14. Bernoulli Naive Bayes UBT	Uni + Bigrams + Trigrams	0.8741	0.8614
15. Logistic Regression UB / \bar{IG}	Uni + Bigrams + Trigrams , $IG \geq \bar{IG}$	0.95	0.9455
16. Logistic Regression UBT / \bar{IG}	Uni + Bigrams + Trigrams , $IG \geq \bar{IG}$	0.95	0.9409

เมื่อนำมาแสดงเป็นคอนฟิวชันเมตริกซ์จะได้ตามตารางที่ 4.2 โดยในแถวเป็นคลาสที่เป็นคำตอบจริง (actual) และในแนวคอลัมน์จะเป็นคลาสที่ทำนายออกมาได้ (predicted)

ตาราง 4.2 คอนฟิวชันเมตริกซ์แสดงผลการทดสอบด้วยชุดทดสอบ (Test set) หลังจากเลือกแผนการทดลองที่มอบอัตราความแม่นยำสูงสุด

	predicted	พอใจ (Positive)	ไม่พอใจ (Positive)
actual			
พอใจ (Positive)		213	7
ไม่พอใจ (Positive)		19	201

สำหรับคอนฟิวชันเมตริกซ์ที่ปรากฏในตารางที่ 4.2 พบว่าการทำนายคลาสที่เป็นความรู้สึกไม่พึงพอใจสามารถทำได้ค่อนข้างดีมากกว่าความรู้สึกที่พึงพอใจ แต่ในคลาสของความรู้สึกที่พึงพอใจสามารถเก็บคลาสจริงของตนเองออกมาได้มากกว่า สำหรับการนำไปใช้งานจริงตัวโมเดลยังไม่ได้ประสิทธิภาพตามที่คาดหวังไว้ ซึ่งเป็นผลมาจากจำนวนประโยคที่ใช้ฝึกสอนยังมีจำนวนที่น้อย และยังมีความหลากหลายไม่เพียงพอ ทำให้ยังไม่ครอบคลุมประโยคที่มีการใช้งานจริงทั้งหมด

มหาวิทยาลัยพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

บทที่ 5

บทสรุป

5.1 อุปสรรคและปัญหา

1. ข้อมูลที่ใช้ในการฝึก (Trained set) ยังมีจำนวนไม่มากพอทั้งในคลาสของความรู้สึกพึงพอใจ (Positive sentiment) และคลาสของความรู้สึกไม่พึงพอใจ (Negative sentiment) ซึ่งเป็นผลทำให้เมื่อนำไปใช้งานจริงยังมีประสิทธิภาพไม่ดีเท่าที่ควร
2. การวิเคราะห์ข้อความประชดประชัน ผลลัพธ์ที่ออกมามีความคลาดเคลื่อนไม่ตรงกับความเป็นจริง
3. การแยกประโยคไม่มีประสิทธิภาพเท่าที่ควร ทำให้การพิจารณาแง่มุมยังไม่เป็นไปตามที่ต้องการ
4. ไม่สามารถตรวจจับคำสำคัญที่เกิดจากการรวมกันของคำตั้งแต่ 4 คำขึ้นไปได้ ทำให้การวิเคราะห์ความรู้สึกคลาดเคลื่อนไม่ตรงกับความเป็นจริง

5.2 แนวทางการพัฒนาต่อ

1. เพิ่มจำนวนประโยคที่ใช้การฝึก โมเดลจำแนกความรู้สึกเพิ่มขึ้น เพื่อให้ครอบคลุมทุกประโยคที่ใช้ในเรื่องของผลตอบรับผลิตภัณฑ์จริง ๆ
2. พัฒนาอัลกอริทึมการแยกประโยคให้มีประสิทธิภาพมากยิ่งขึ้น โดยใช้ภาคประธาน และภาคแสดงมารวมพิจารณาในการแยกประโยคด้วย
3. พัฒนาอัลกอริทึม สำหรับการพิจารณาคำที่จะเลือกมาเป็นคำสำคัญในรูปแบบอื่น
4. ทดสอบแยกพีเจอร์ในรูปแบบอื่น ๆ แทนถุงคำ (bag of words)
5. ทดลองนำโมเดลจำแนกความรู้สึก ไปใช้กับกลุ่มผลิตภัณฑ์อื่น ๆ ที่มีสรรพคุณที่มอบให้กับผู้บริโภคใกล้เคียงกับโลชั่นบำรุงผิวกาย เช่น กลุ่มของผลิตภัณฑ์บำรุงผิวหน้า เป็นต้น
6. สำหรับการนำไปใช้งานจริงหากต้องการความถูกต้องแม่นยำเพิ่มขึ้น ต้องเปิดให้ผู้ใช้งานในระบบได้มีส่วนร่วมในการพัฒนาระบบไปพร้อมกันด้วย โดยให้พวกเขาสามารถให้คะแนนสำหรับผลของการทำนายว่าเป็นไปตามความเป็นจริงหรือไม่ และนำข้อมูลที่ผ่านมาการให้คะแนนมาแล้วเหล่านั้น กลับมาฝึกจำแนกความรู้สึกใหม่อีกครั้ง เพื่อพัฒนาความถูกต้องแม่นยำของโมเดลต่อไป

บรรณานุกรม

สำนักงานราชบัณฑิตยสภา. 2558. **หลักเกณฑ์การเว้นวรรค.** [Online]. Available:

http://www.royin.go.th/?page_id=629

สมบูรณ์ พัฒน์ธีรพงศ์. 2554. **JavaScript Programming Guide.** กรุงเทพฯ.

คอนเทนต์บลู พับลิชซิ่ง.

กานดา รุณนะพงศา และปโยธร อุราธรรมกุล. 2549. **การตัดคำภาษาไทยโดยการปรับปรุงกฎ**

และพจนานุกรมแบบใหม่. รายงานวิจัย ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะ
วิศวกรรมศาสตร์, มหาวิทยาลัยขอนแก่น.

Virach Sornlertlamvanich, Thatsanee Charoenporn & Hitoshi. 2556.

ORCHID: Thai Part-Of-Speech Tagged Corpus.

Wannaphong Phatthiyaphaibun. 2560. **PyThaiNLP documentation.** [Online]. Available:

<https://thainlp.org/pythainlp/docs/1.7/>

Python Software Foundation US. 2018. **Python 3.5.6 documentation.** [Online] Available:

<https://docs.python.org/3.5/>

Google Inc. 2019. **Firestore documentation.** [Online] Available:

<https://firebase.google.com/docs>

Scrapinghub Ltd. 2019. **Scrapy 1.6 documentation.** [Online] Available:

<https://docs.scrapy.org/en/latest/>

Inria. 2019. **Documentation of scikit-learn.** [Online] Available:

<https://scikit-learn.org/stable/documentation.html>

Facebook Inc. 2019. **React documentation**. [Online] Available:

<https://reactjs.org/docs/getting-started.html>

Leonard Richardson. 2019. **Beautiful Soup Documentation**. [Online] Available:

<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

