การพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษาบนเฟซบุ๊ก โดยใช้ตัวจำแนกข้อมูล นาอีฟ เบย์ สำหรับภาษาไทย

นายสมัคร ชัยสงวน

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
แขนงวิชาเทคโนโลยีสารสนเทศและการสื่อสาร สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
พ.ศ. 2561

Development of a Real-time Sentiment Analysis System of Students on Facebook Using Naive Bayes Classifier in Thai Language

Mr.Samark Chaisanguan

A Thesis Submitted in Partial Fulfillment of the Requirements for

the Degree of Master of Science in Information and Communication Technology

School of Science and Technology

Sukhothai Thammathirat Open University

2018

หัวข้อวิทยานิพนธ์	การพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูล นาอีฟ เบย์ สำหรับภาษาไทย
ชื่อและนามสกุล	นายสมัคร ชัยสงวน
แขนงวิชา	เทคโนโลยีสารสนเทศและการสื่อสาร
สาขาวิชา	วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
อาจารย์ที่ปรึกษา	1. ผู้ช่วยศาสตราจารย์ ดร.วฤษาย์ ร่มสายหยุด 2. ผู้ช่วยศาสตราจารย์ ดร.ปราโมทย์ ลือนาม

วิทยานิพนธ์นี้ ได้รับความเห็นชอบให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรระดับปริญญาโท เมื่อวันที่ 19 ตุลาคม 2561

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(อาจารย์ ดร.ธงชัย แก้วกิริยา)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.วฤษาย์ ร่มสายหยุด)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ปราโมทย์ ลือนาม)

.....ประธานกรรมการบัณฑิตศึกษา
(รองศาสตราจารย์ ดร.กฤษณา รุ่งโรจน์วณิชย์)

ชื่อวิทยานิพนธ์ การพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูล นาอีฟ เบย์ สำหรับภาษาไทย

ผู้วิจัย นายสมัคร ชัยสงวน **รหัสนักศึกษา** 2599600422

ปริญญา วิทยาศาสตรมหาบัณฑิต (เทคโนโลยีสารสนเทศและการสื่อสาร)

อาจารย์ที่ปรึกษา (1) ผู้ช่วยศาสตราจารย์ ดร.วฤษาย์ ร่มสายหยุด

(2) ผู้ช่วยศาสตราจารย์ ดร.ปราโมทย์ ลือนาม **ปีการศึกษา** 2561

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ สำหรับภาษาไทย และ 2) เพื่อประเมินมาตรการความถูกต้องของค่าความแม่นยำและค่าเรียกคืนของระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ สำหรับภาษาไทย

งานวิจัยนี้รวบรวมข้อมูลจากเฟซบุ๊กเพจของมหาวิทยาลัยสุโขทัยธรรมาธิราช ซึ่งเป็นข้อมูลการแสดงความรู้สึกของนักศึกษา ที่ติดตามเฟซบุ๊กเพจของมหาวิทยาลัย เพื่อทำการวิเคราะห์ความรู้สึกของนักศึกษาแบบเรียลไทม์ด้วยนาอีฟ เบย์ จากหลักการพื้นฐานของการประมวลผลภาษาธรรมชาติด้วยวิธีการตัดคำภาษาไทย และประยุกต์ตัวจำแนกข้อมูลนาอีฟ เบย์ โดยผลลัพธ์ทั้งหมดสามารถแสดงผลแบบเรียลไทม์ กับระบบจัดการคำคลังคำศัพท์ สรุปผลการวิเคราะห์ความรู้สึก เครื่องมือทดสอบประโยค และระบบแจ้งเตือนข้อความผ่านทางไลน์

ผลการวิจัยนี้ได้ผลลัพธ์ค่าความแม่นยำร้อยละ 80.03 ค่าเรียกคืนร้อยละ 90.52 ค่าความถูกต้อง 89.63 และการวัดประสิทธิภาพโดยรวมร้อยละ 84.81 ตามลำดับ

คำสำคัญ ระบบวิเคราะห์ความรู้สึก ตัวจำแนกข้อมูลนาอีฟ เบย์ ระบบเรียลไทม์ เฟซบุ๊กเพจของมหาวิทยาลัยสุโขทัยธรรมาธิราช

Thesis title: Development of a Real-time Sentiment Analysis System of Students on Facebook Using Naive Bayes Classifier in Thai Language

Researcher: Mr.Samark Chaisanguan; **ID**: 2599600422;

Degree: Master of Science (Information and Communication Technology);

Thesis advisors: (1) Dr.Walisa Romsaiyud, Assistant Professor;

(2) Dr.Pramote Luenam, Assistant Professor; **Academic year**: 2018

Abstract

The purposes of this research were as follows: 1) to develop a real-time sentiment analysis system of students on Facebook using Naive Bayes classifier in Thai language, and 2) to evaluate the accuracy of precision and recall measures of real-time sentiment analysis system of students on Facebook using Naive Bayes classifier in Thai language.

The research collected data from the Facebook page of Sukhothai Thammathirat Open University which comprise the sentiment expression of students who followed the university's Facebook fan page. This method was employed as the means to perform an analysis of students' a real-time sentiment analysis using Naive Bayes. From the fundamental principles of Natural Language Processing with a Thai word segmentation method and applied Naive Bayes Classifier. All results were able to demonstrate in real-time results of the vocabulary management system, the results of sentiment analysis, the tool for testing sentences, and Line notification system.

The measurement results were 80.03% for precision, 90.52% for recall, 89.63% for accuracy, and 84.81% for F-measure respectively.

Keywords: Sentiment analysis system, Naive Bayes Classifier, Real-time system, STOU Facebook fan page

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่ง ผู้วิจัยขอขอบพระคุณผู้ช่วยศาสตราจารย์ ดร.วฤษาย์ ร่มสายหยุด อาจารย์ที่ปรึกษาหลัก และผู้ช่วยศาสตราจารย์ ดร.ปราโมทย์ ลือนาม อาจารย์ที่ปรึกษาร่วม ที่ได้มอบความรู้คำแนะนำที่เป็นประโยชน์ต่องานวิจัย ให้การสนับสนุนและช่วยเหลือในการแก้ไขปัญหา ตลอดจนข้อมูลที่จำเป็นและแนวทางในการทำวิจัย ท่านช่วยแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ พร้อมทั้งให้กำลังใจ ผู้วิจัยรู้สึกซาบซึ้งในความกรุณาของท่านและขอขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอขอบพระคุณคณาจารย์สาขาวิชาวิทยาศาสตร์และเทคโนโลยี แขนงเทคโนโลยีสารสนเทศและการสื่อสาร ทุกท่านที่ได้ถ่ายทอดความรู้ซึ่งเป็นประโยชน์แก่ผู้ทำวิจัย และทำให้วิทยานิพนธ์ฉบับนี้สำเร็จได้ และขอขอบคุณบัณฑิตวิทยาลัย มหาวิทยาลัยสุโขทัยธรรมาธิราชที่ได้ให้คำแนะนำต่าง ๆ ในขั้นตอนของการดำเนินการวิจัย

ขอขอบคุณพี่ ๆ น้อง ๆ นักศึกษาปริญญาโททุกคนที่ให้คำแนะนำ แนวทางในการทำงานวิจัยคอยช่วยเหลือ และให้กำลังใจในการทำวิทยานิพนธ์ด้วยดีตลอดมา และผู้มีพระคุณทุกท่านที่มิได้เอ่ยนามอันทำให้วิทยานิพนธ์ฉบับนี้สำเร็จไปได้ด้วยดี

ประโยชน์อันใดอันจะพึงเกิดขึ้นจากวิทยานิพนธ์ฉบับนี้ผู้วิจัยขอมอบแก่บิดามารดาที่ให้การสนับสนุนและช่วยเหลือ อีกทั้งคอยเป็นกำลังใจตลอดมา

สมัคร ชัยสงวน

ตุลาคม 2561

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....ง

บทคัดย่อภาษาอังกฤษ.....จ

กิตติกรรมประกาศ.....ฉ

สารบัญตาราง.....ฌ

สารบัญภาพ.....ฎ

บทที่ 1 บทนำ.....1

1.1 ความเป็นมาและความสำคัญของปัญหา.....1

1.2 วัตถุประสงค์ของการวิจัย.....3

1.3 กรอบแนวคิดการวิจัย.....3

1.4 ขอบเขตงานวิจัย.....4

1.5 ประโยชน์ที่คาดว่าจะได้รับ.....6

1.6 นิยามศัพท์เฉพาะ.....7

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....8

2.1 ลักษณะของภาษาไทย.....8

2.2 การสืบค้นข้อมูลเฟซบุ๊กด้วย Facebook Graph API.....12

2.3 การตัดคำภาษาไทย (Thai Word Segmentation).....16

2.4 การวิเคราะห์ความรู้สึก (Sentiment Analysis).....24

2.5 ตัวจำแนกข้อมูลนาอีฟ เบย์ (Naive Bayes Classifier).....25

2.6 การแจ้งเตือนผ่านทาง Line ด้วย Line Messaging API.....27

2.7 ระบบเรียลไทม์ (Real-time System).....31

2.8 การทดสอบประสิทธิภาพ (Performance Evaluation).....34

2.9 งานวิจัยที่เกี่ยวข้อง.....36

บทที่ 3 วิธีการดำเนินงานวิจัย.....41

3.1 การศึกษาและวิเคราะห์ปัญหา.....41

สารบัญ (ต่อ)

หน้า

3.2 การวิเคราะห์และออกแบบขั้นตอนการทำงานของระบบ....42
3.3 ข้อมูลและกลุ่มตัวอย่างที่ใช้ในงานวิจัย....72
3.4 การพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ (Real-time) 78
3.5 การทดสอบประสิทธิภาพของระบบ....80
บทที่ 4 ผลการดำเนินงาน 83
4.1 ผลการพัฒนาระบบ....83
4.2. ผลการประเมินประสิทธิภาพของระบบ....95
บทที่ 5 สรุปการวิจัย อภิปรายผล และข้อเสนอแนะ....99
5.1 สรุปการวิจัย....99
5.2 ปัญหาและอุปสรรค....101
5.3. ข้อเสนอแนะ....102
บรรณานุกรม....103
ประวัติผู้วิจัย....107

สารบัญตาราง

หน้า

ตารางที่ 2.1 พยัญชนะไทย 44 รูป....8
ตารางที่ 2.2 สระ 21 รูป....9
ตารางที่ 2.3 ประเภทของ Access Token Facebook....14
ตารางที่ 2.4 ตัวอย่างการแบ่งคำด้วยวิธีการเทียบคำที่ยาวที่สุด (Longest Matching) 18
ตารางที่ 2.5 เปรียบเทียบหลักการตัดคำ....22
ตารางที่ 2.6 สรุปคุณลักษณะข้อเด่น ข้อด้อย และตัวอย่างการนำไปใช้งานที่เหมาะสม....22
ตารางที่ 2.7 อัตราการใช้บริการส่งข้อความด้วย LINE API....29
ตารางที่ 2.8 ข้อมูลการตอบกลับจากเซิฟเวอร์ไลน์....30
ตารางที่ 2.9 ข้อมูล Header การตอบกลับจากเซิฟเวอร์ไลน์....30
ตารางที่ 2.10 ข้อมูลสถานการณ์ตอบกลับในกรณีที่เกิดข้อผิดพลาดจากเซิฟเวอร์ไลน์....31
ตารางที่ 3.1 ตัวอย่างข้อมูลที่ใช้ในการเรียกใช้งาน Facebook Graph API....44
ตารางที่ 3.2 ประเภทของคำในพจนานุกรม....48
ตารางที่ 3.3 คำอธิบายยูสเคส ตรวจสอบผู้ใช้งาน....68
ตารางที่ 3.4 คำอธิบายยูสเคส จัดการคำศัพท์....68
ตารางที่ 3.5 คำอธิบายยูสเคส ทดสอบประโยค....69
ตารางที่ 3.6 คำอธิบายยูสเคส รายงานข้อความผู้ใช้งานเฟซบุ๊ก....70
ตารางที่ 3.7 คำอธิบายยูสเคส ข้อความแจ้งเตือน....70
ตารางที่ 3.8 คำอธิบายยูสเคส ผลสรุปวิเคราะห์ความรู้สึก....71
ตารางที่ 3.9 รายละเอียดการจัดเก็บข้อมูล....72
ตารางที่ 3.10 ตัวอย่างข้อมูลการแสดงความคิดเห็น....72
ตารางที่ 3.11 รายละเอียดการจัดเก็บข้อมูลคลังคำศัพท์....73
ตารางที่ 3.12 ตัวอย่างข้อมูลคลังคำศัพท์ที่....73
ตารางที่ 3.13 รายละเอียดการจัดเก็บฐานข้อมูลคลังคำศัพท์แยกตามประเภทความรู้สึก....74
ตารางที่ 3.16 ตัวอย่างข้อมูลคลังคำศัพท์ ประเภทความรู้สึกเป็นกลาง....76

สารบัญตาราง (ต่อ)

หน้า

ตารางที่ 3.17 รายละเอียดการจัดเก็บฐานข้อมูลความคิดเห็นที่ผ่านกระบวนการตัดคำภาษาไทย....77
ตารางที่ 3.18 ตัวอย่างข้อมูลการตัดคำและผลลัพธ์การวิเคราะห์ความรู้สึก.....77
ตารางที่ 3.19 รายละเอียดการจัดเก็บฐานข้อมูลผู้ดูแลระบบ.....78
ตารางที่ 3.20 ตัวอย่างข้อมูลผู้ดูแลระบบที่จัดเก็บไว้ในฐานข้อมูลระบบ.....78
ตารางที่ 4.1 ผลลัพธ์การตัดคำภาษาไทย.....92
ตารางที่ 4.2 ผลลัพธ์การวิเคราะห์ความรู้สึก.....92
ตารางที่ 4.3 สรุปข้อมูลที่ใช้งานระบบ.....93
ตารางที่ 4.4 ผลการวิเคราะห์ความรู้สึกเชิงบวก (Positive) 5 อันดับแรก.....94
ตารางที่ 4.5 ผลการวิเคราะห์ความรู้สึกเชิงลบ (Negative) 5 อันดับแรก.....94
ตารางที่ 4.6 ผลการวิเคราะห์คำถามที่พบบ่อย 5 อันดับแรก.....95
ตารางที่ 4.7 ผลการทดสอบประสิทธิภาพรอบที่ 1.....96
ตารางที่ 4.8 ผลการทดสอบประสิทธิภาพรอบที่ 2.....96
ตารางที่ 4.9 ผลการทดสอบประสิทธิภาพรอบที่ 3.....97
ตารางที่ 4.10 ผลการทดสอบประสิทธิภาพรอบที่ 4.....97
ตารางที่ 4.11 ผลการทดสอบประสิทธิภาพรอบที่ 5.....98
ตารางที่ 4.12 สรุปผลการทดสอบประสิทธิภาพของระบบ.....98

สารบัญภาพ

หน้า

ภาพที่ 1.1 กรอบแนวคิดภาพรวมของงานวิจัย.....3
ภาพที่ 2.1 ตัวอย่างข้อมูล JSON ของเฟซบุ๊กเพจ.....15
ภาพที่ 2.2 ตัวอย่างการแบ่งคำด้วยวิธีการย้อนกลับ (Backtracking) 19
ภาพที่ 2.3 ตัวอย่างโค้ด Websocket.....33
ภาพที่ 3.1 กระบวนการทำงานโดยรวมของระบบ.....42
ภาพที่ 3.2 กระบวนการทำงานของระบบในการรวบรวมข้อมูล.....43
ภาพที่ 3.3 ตัวอย่าง URL และชุดข้อมูลที่ใช้ในการเรียกใช้ข้อมูล Facebook Graph API.....45
ภาพที่ 3.4 ตัวอย่างข้อมูลการตอบกลับจาก Facebook Graph API กรณีเกิดข้อผิดพลาด.....45
ภาพที่ 3.5 ข้อมูลความคิดเห็นจากเฟซบุ๊กในรูปแบบ JSON.....46
ภาพที่ 3.6 ตัวอย่างโค้ดการถอดรหัสข้อมูลจาก JSON เป็นข้อมูลชนิดอาร์เรย์.....47
ภาพที่ 3.7 ขั้นตอนกระบวนการตัดคำออกจากข้อความภาษาไทย.....48
ภาพที่ 3.8 ขั้นตอนกระบวนวิเคราะห์ความรู้สึกจากข้อความ.....50
ภาพที่ 3.9 แผนผังงานระบบ (Context Diagram).....51
ภาพที่ 3.10 แผนภาพการไหลของข้อมูลระดับที่ 1 (Dataflow Diagram Level 1).....52
ภาพที่ 3.11 กระบวนการจัดการการเข้าถึงข้อมูล (Authentication).....53
ภาพที่ 3.12 กระบวนดูรายงานสถิติต่าง ๆ (Dashboard Viewer).....54
ภาพที่ 3.13 กระบวนการจัดการข้อมูลคลังคำศัพท์ (Dictionary Management).....55
ภาพที่ 3.14 กระบวนการค้นหาคลังคำศัพท์ (Search Corpus).....56
ภาพที่ 3.15 กระบวนการอัพโหลดคำศัพท์ (Upload Corpus).....57
ภาพที่ 3.16 กระบวนการสร้างคำศัพท์ใหม่ (Create Corpus).....58
ภาพที่ 3.17 กระบวนการสร้างคำศัพท์ใหม่ (Create Corpus).....59
ภาพที่ 3.18 กระบวนการออกรายงานผลการวิเคราะห์ความคิดเห็น (Report).....60
ภาพที่ 3.19 กระบวนการสมัครรับการแจ้งเตือนผ่านทาง (Subscribe Line Notification).....61
ภาพที่ 3.20 กระบวนการทดสอบข้อความ (Test Sentences).....62

สารบัญภาพ (ต่อ)

หน้า

ภาพที่ 3.21 กระบวนการตัดคำ (Word segmentation).....63
ภาพที่ 3.22 กระบวนการวิเคราะห์ความรู้สึก (Sentiment Analysis).....64
ภาพที่ 3.23 กระบวนการให้บริการข้อมูล (Data Provider).....65
ภาพที่ 3.24 กระบวนการทำงานเบื้องหลัง (Background Process).....66
ภาพที่ 3.25 ยูสเคสไดอะแกรมสำหรับการพัฒนาระบบ.....67
ภาพที่ 3.26 การประเมินผลโดยวิธี K-fold Cross-validation.....81
ภาพที่ 4.1 หน้าจอแสดงผลของระบบ 84
ภาพที่ 4.2 หน้าจอตรวจสอบผู้ใช้งาน.....85
ภาพที่ 4.3 การแสดงข้อความเมื่อไม่พบข้อมูลผู้ดูแลระบบในระบบ.....85
ภาพที่ 4.4 ข้อความแจ้งเตือนกรณีไม่ได้กรอกบัญชีผู้ใช้งาน.....86
ภาพที่ 4.5 หน้าจอแสดงการออกจากระบบ.....87
ภาพที่ 4.6 หน้าจอจัดการข้อมูลระบบ.....88
ภาพที่ 4.7 หน้าจอแสดงผลส่วนจัดการคำศัพท์ของระบบ.....88
ภาพที่ 4.8 หน้าจอแสดงผลส่วนจัดการเพิ่มคำศัพท์ในพจนานุกรมของระบบ.....89
ภาพที่ 4.9 ส่วนรายงานข้อความผู้ใช้งานเฟซบุ๊ก.....89
ภาพที่ 4.10 หน้าจอสำหรับอัพโหลดพจนานุกรม.....90
ภาพที่ 4.11 หน้าจอสำหรับทดสอบประโยค.....90
ภาพที่ 4.12 คิวอาร์โค้ดรับการแจ้งเตือน.....91
ภาพที่ 4.13 ข้อความแจ้งเตือน Line Notification.....91

บทที่ 1
บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

เฟซบุ๊กเป็นโซเชียลมีเดียยอดนิยมอันดับ 1 ของคนไทย (wp, 2018) โดยมีจำนวนผู้ใช้งานอยู่ที่ 51 ล้านคน คิดเป็นร้อยละ 90 ซึ่งจากสถิติประเทศไทยยังเป็นประเทศที่ใช้เวลาต่อวันอยู่กับอินเทอร์เน็ตมากที่สุดในโลกโดยเฉลี่ยอยู่ที่ 9 ชั่วโมง 38 นาทีต่อวัน และใช้เวลาในการเล่นโซเชียลมีเดียโดยเฉลี่ยอยู่ที่ 3 ชั่วโมง 10 นาทีต่อวัน ด้วยเฟซบุ๊กถือว่าเป็นเครื่องมือที่นิยมใช้ในการติดต่อสื่อสาร และเป็นแหล่งแสดงความรู้สึกของผู้คนจำนวนมาก การนำข้อมูลจากเพจเฟซบุ๊กมาใช้ในการวิเคราะห์ความรู้สึก (Sentiment Analysis) ของผู้ติดตามเพจในเฟซบุ๊กนั้นถือว่าสำคัญโดยเฉพาะในด้านของการประสบความสำเร็จทางธุรกิจเพราะ “ลูกค้า” คือปัจจัยสำคัญที่สุดที่จะทำให้ธุรกิจประสบความสำเร็จ การสร้างความสัมพันธ์ที่ดีเพื่อรักษาฐานข้อมูลลูกค้าคือการ “รับความรู้สึกของลูกค้า” เป็นกระบวนการทางธุรกิจในการเข้าใจลูกค้า เพื่อใช้ประโยชน์ในการพัฒนาสินค้าและบริการ โดยส่งเสริมให้เกิดการจงรักภักดีต่อสินค้าและผลิตภัณฑ์ นอกจากนี้หน่วยงานภาครัฐสามารถนำความรู้ไปดำเนินการประยุกต์ในด้าน “ความรู้สึกของประชาชน” กับหลายๆ หน่วยงานภาครัฐ ซึ่งเป็นผลให้ข้อมูลมีปริมาณมากและเพิ่มขึ้นอยู่ตลอดเวลาอย่างรวดเร็ว เป็นเรื่องยากที่จะรวบรวมข้อมูลทั้งหมดมาอ่านหรือวิเคราะห์โดยการใช้บุคลากร จึงจำเป็นต้องนำระบบคอมพิวเตอร์เข้ามาช่วยในการแก้ปัญหาเพื่อเพิ่มความสามารถในการแข่งขัน

ในปัจจุบันการศึกษาระดับอุดมศึกษามีการแข่งขันกันมากขึ้นผู้เรียนมีทางเลือกในการตัดสินใจเลือกสถานศึกษามากขึ้นมหาวิทยาลัยจึงจำเป็นอย่างยิ่งที่จะต้องมีกลยุทธ์ในการหาวิธีการที่จะให้กลุ่มเป้าหมายตัดสินใจเลือกเข้าศึกษาต่อในมหาวิทยาลัยโดยการจัดกิจกรรมต่าง ๆ และรูปแบบการประชาสัมพันธ์ที่หลากหลายช่องทาง มหาวิทยาลัยจำเป็นต้องสร้างความพึงพอใจแก่ผู้เรียนหรือผู้มารับบริการด้วย มีความจำเป็นอย่างยิ่งที่จะต้องรับรู้และเข้าใจความรู้สึกของนักศึกษาที่มีต่อมหาวิทยาลัย เพื่อมหาวิทยาลัยจะได้นำไปใช้เป็นแนวทางในการปรับปรุงการจัดการเรียนการสอน และพัฒนาระบบการบริการของมหาวิทยาลัยต่อไป มหาวิทยาลัยสุโขทัยธรรมาธิราชเป็นหนึ่งในมหาวิทยาลัยในประเทศไทย ที่มีการใช้เฟซบุ๊กเป็นช่องทางในการติดต่อสื่อสารกับนักศึกษา โดยใช้ชื่อว่า STOUniversity (https://fb.com/STOUniversity/) มีนักศึกษาที่ติดตามเฟซบุ๊กเพจของ

มหาวิทยาลัยจำนวน 241,935 คน (สืบค้นเมื่อ 6 กุมภาพันธ์ 2561, จาก https://fb.com/STOUniversity/) ซึ่งมีหลากหลายความความคิดเห็นที่ส่งเข้ามาในเฟซบุ๊กเพจ ตัวอย่างความคิดเห็นเชิงบวกเช่น “ภูมิใจที่ได้เรียนมสธ” ความคิดเห็นเชิงลบเช่น “ข้อสอบยากมาก” และความคิดเห็นที่เป็นกลาง “แชร์ได้ไหมครับท่านอาจารย์”

ภาษาไทยเป็นภาษาที่มีลักษณะเฉพาะทางโครงสร้างที่ซับซ้อน ตั้งแต่โครงสร้างตัวอักษร ที่มีพยัญชนะ 44 รูป สระ 21 รูป วรรณยุกต์ 4 รูป ตัวเลข 10 รูป และเครื่องหมายวรรคตอน พร้อมทั้งสัญลักษณ์พิเศษต่าง ๆ เช่น ไม้ทัณฑฆาต จุลภาค และมหัพภาค เป็นต้น ภาษาไทยจึงเป็นภาษาที่ยากเมื่อเทียบกับภาษาอื่น เพราะเป็นภาษาที่ไม่มีการเขียนแบ่งพยางค์ คำ วลี หรือประโยค ไม่มีหลักเกณฑ์ตายตัว การสะกดคำมี กฎ ไวยากรณ์ ที่มีรูปแบบซับซ้อน มีคำยืม คำทับศัพท์ คำเฉพาะ คำเกิดใหม่จำนวนมาก และรวมถึงลักษณะของคำที่มีความกำกวมสูง จึงถือได้ว่า ประโยคในภาษาไทยมีความซับซ้อน ทำให้บางครั้งการตีความหมายของประโยคไม่เหมือนกัน แต่มีความหมายเหมือนกัน เช่น คำกำกวม ที่มีความหมายไม่แน่ชัด และภาษาไทยเป็นภาษาที่เรียงคำ การเรียงคำในภาษาไทยเป็นเรื่องที่สำคัญมาก เพราะการเปลี่ยนตำแหน่งของคำ ทำให้ความหมายของคำเปลี่ยนไปด้วย ซึ่งภาษาไทยไม่มีสัญลักษณ์ที่สามารถบ่งบอกถึงขอบเขตของคำ ที่เรียงต่อเนื่องกัน ทั้งประโยคเหมือนกับภาษาอังกฤษที่ใช้ช่องว่าง (Space) คั่นระหว่างขอบเขตของคำ เนื่องจากภาษาไทยเป็นภาษาที่ซับซ้อนจึงทำให้โปรแกรมคอมพิวเตอร์รู้จักคำในภาษาไทยนั้นมีความยุ่งยากกว่าในภาษาอังกฤษ โดยเฉพาะประโยคส่วนใหญ่บนอินเทอร์เน็ต ที่นิยมใช้ภาษาที่ไม่เป็นทางการหรือภาษาพูด รวมทั้งมีโครงสร้างประโยคที่ไม่ถูกต้องตามหลักไวยากรณ์ภาษาไทย ยิ่งทำให้ยากต่อการนำมาวิเคราะห์

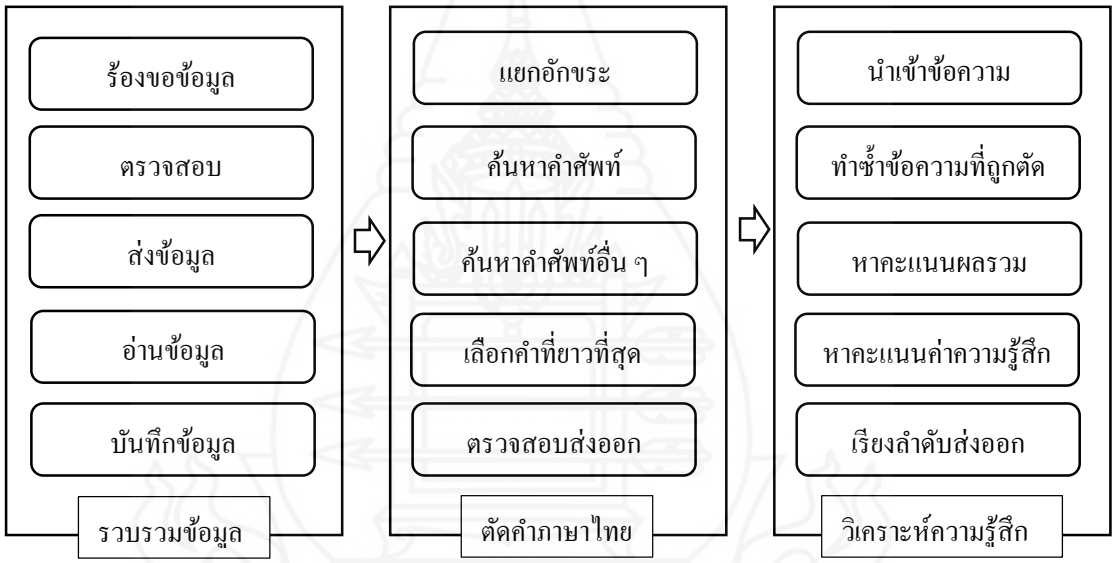
งานวิจัยส่วนใหญ่มุ่งเน้นในการพัฒนาระบบการวิเคราะห์ความรู้สึกที่เป็นข้อความภาษาไทย ที่วิเคราะห์ความรู้สึกจากเฟซบุ๊กเพจที่เป็นข้อความภาษาไทยได้อย่างมีประสิทธิภาพและสามารถวิเคราะห์ได้แบบเรียลไทม์ งานวิจัยนี้จึงนำเสนอการพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ (Real-time) ของนักศึกษาบน เฟซบุ๊กโดยใช้ตัวจำแนกข้อมูล นาอีฟ เบย์ สำหรับภาษาไทย ซึ่งเป็นวิธีจำแนกประเภทข้อมูลที่มีประสิทธิภาพวิธีหนึ่ง ซึ่งมีการนำไปประยุกต์ใช้งานในด้านการจำแนกประเภทข้อความ ที่ใช้งานได้ดีไม่ต่างจากการจำแนกประเภทวิธีการอื่น เนื่องจากเป็นวิธีการจำแนกข้อมูลได้อย่างมีประสิทธิภาพและอัลกอริทึมในการทำงานที่ไม่ซับซ้อน โดยผู้วิจัยจะรวบรวมข้อมูลโดยใช้ Facebook Graph API ร่วมกับภาษา PHP และภาษา Java Script (Socket.IO) มาเก็บไว้ที่ฐานข้อมูล เพื่อนำมาวิเคราะห์ความเห็นออกมาเป็นคะแนนบวก (+1) ลบ (-1) หรือเป็นกลาง (0) เพื่อวิเคราะห์ความรู้สึก (Sentiment Analysis) ของผู้ติดตามเพจที่เป็นเชิงบวก (Positive) และเชิงลบ (Negative)

2. วัตถุประสงค์ของการวิจัย

2.1 เพื่อพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษาบนเฟซบุ๊ก โดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ สำหรับภาษาไทย

2.2 เพื่อประเมินมาตรการความถูกต้องของค่าความแม่นยำและค่าเรียกคืนของระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ สำหรับภาษาไทย

3. กรอบแนวคิดการวิจัย



ภาพที่ 1.1 กรอบแนวคิดภาพรวมของงานวิจัย

จากภาพที่ 1.1 แสดงกรอบแนวคิดภาพรวมของงานวิจัย ซึ่งมี 3 ขั้นตอนด้วยกันคือ

ขั้นตอนที่ 1 เรียกใช้ข้อมูล เป็นการรวบรวมข้อมูลจากเฟซบุ๊ก (Facebook) โดยการส่งคำขอข้อมูลไปยังเฟซบุ๊ก เมื่อเฟซบุ๊กอนุมัติ ระบบจะดึงข้อมูลจากเฟซบุ๊กเพจมหาวิทยาลัยสุโขทัยธรรมาธิราช STOUniversity (https://fb.com/STOUniversity/) มาเก็บไว้ที่ฐานข้อมูลของระบบ แล้วส่งข้อมูลไปยังขั้นตอนการตัดคำภาษาไทย

ขั้นตอนที่ 2 ตัดคำภาษาไทย หลังจากที่ได้ข้อมูลความคิดเห็นบนเฟซบุ๊กแล้ว จะนำเข้าสู่กระบวนการตัดคำภาษาไทย โดยใช้วิธีการเทียบคำที่ยาวที่สุด (Longest Word Pattern

Matching) จากการพิจารณาด้วยวิธีวนลูปจากซ้ายไปขวาตามหลักภาษาไทยเพื่อเทียบกับพจนานุกรม Lexicon – Thai จากนั้นจะส่งข้อมูลไปยังกระบวนการวิเคราะห์ความรู้สึก

ขั้นตอนที่ 3 วิเคราะห์ความรู้สึก นำเข้าข้อความที่ถูกตัดจากกระบวนการตัดคำแล้วนำเข้าสู่กระบวนการวิเคราะห์ความรู้สึก (Sentiment Analysis) โดยการวิเคราะห์ความเห็นออกมาเป็นคะแนนบวก (+1) ลบ (-1) หรือเป็นกลาง (0) เพื่อสรุปภาพรวมการวิเคราะห์ความรู้สึก (Sentiment Analysis) แบบ 2 กลุ่ม คือ ความรู้สึกเชิงบวก (Positive) และความรู้สึกเชิงลบ (Negative) เพื่อจัดเก็บไว้ในฐานข้อมูลและแสดงผลผ่านหน้าจอของระบบ

4. ขอบเขตงานวิจัย

4.1 ขอบเขตด้านข้อมูล ระบบวิเคราะห์ความรู้สึกแบบทเรียลไทม์ (Real-time) ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ (Naive Bayes Classifier) สำหรับภาษาไทยมาทำการจำแนกข้อมูลใช้ข้อมูลจากเฟซบุ๊กเพจมหาวิทยาลัยสุโขทัยธรรมาธิราช STOUniversity (https://fb.com/STOUniversity/) ซึ่งเป็นข้อมูลการแสดงความคิดเห็นของนักศึกษาที่ติดตามเฟซบุ๊กเพจของมหาวิทยาลัย จำนวน 241,935 คน

4.2 ขอบเขตด้านระบบงาน โดยพัฒนาระบบในรูปแบบของเว็บไซต์ (Web Site) ด้วยภาษา PHP บน Laravel Framework 5.3 ร่วมกับ Mysql ทำงานในส่วนของการจัดเก็บฐานข้อมูลออนไลน์โดยแบ่งออกเป็นสองส่วนหลักคือส่วนของผู้ใช้งาน และส่วนของผู้ดูแลระบบดังต่อไปนี้

4.2.1 ส่วนของผู้ใช้งาน

1) สามารถแสดงจำนวนผลการวิเคราะห์ความคิดเห็นเชิงบวกได้
2) สามารถแสดงจำนวนผลการวิเคราะห์ความคิดเห็นเชิงลบได้
3) สามารถแสดงจำนวนผู้กดติดตาม (Follow) เฟซบุ๊กเพจได้
4) สามารถแสดงจำนวนผู้กดไลค์ (Like) เฟซบุ๊กเพจได้
5) สามารถแสดงผลการวิเคราะห์เชิงบวกได้
6) สามารถแสดงผลการวิเคราะห์เชิงลบได้

4.2.2 ส่วนของผู้ดูแลระบบ

1) ระบบตรวจสอบการใช้งาน

(1) สามารถตรวจสอบการเข้าใช้งานระบบได้
(2) สามารถตรวจสอบระยะเวลาหมดอายุในการใช้งานระบบได้
(3) สามารถออกจากระบบได้

2) ระบบแผงควบคุม (Dash Board)

(1) สามารถแสดงข้อมูลจำนวนผู้กดติดตามเฟซบุ๊กเพจได้

(2) สามารถแสดงข้อมูลจำนวนผู้กดไลค์เฟซบุ๊กเพจได้

(3) สามารถแสดงจำนวนข้อมูลคลังคำศัพท์ได้

(4) สามารถแสดงจำนวนข้อมูลคลังคำศัพท์ประเภทความรู้สึกได้

(5) สามารถแสดงจำนวนผลการวิเคราะห์ความคิดเห็นเชิงบวกได้

(6) สามารถแสดงจำนวนผลการวิเคราะห์ความคิดเห็นเชิงลบได้

(7) สามารถแสดงจำนวนผลการวิเคราะห์ความคิดเห็นเป็นกลางได้

3) ระบบจัดการคลังคำศัพท์

(1) สามารถเพิ่มข้อมูลคลังคำศัพท์ได้

(2) สามารถแก้ไขข้อมูลคลังคำศัพท์ได้

(3) สามารถลบข้อมูลคลังคำศัพท์ได้

(4) สามารถค้นหาคลังคำศัพท์ได้

4) ระบบทดสอบประโยค

(1) สามารถนำเข้าข้อความมาทดสอบได้

(2) สามารถดูผลการทดสอบได้

(3) สามารถสร้างคำศัพท์ใหม่ผ่านเครื่องมือสร้างคำศัพท์ได้

(4) สามารถดูคะแนนผลการวิเคราะห์ข้อความได้

(5) สามารถแก้ไขข้อมูลคำศัพท์ได้

5) ระบบรายงานการวิเคราะห์ความรู้สึกของผู้ใช้งานเฟซบุ๊ก

(1) สามารถดูรายงานการวิเคราะห์ความรู้สึกได้

(2) สามารถค้นหาข้อมูลความคิดเห็นได้

(3) สามารถค้นหาข้อมูลความคิดเห็นเชิงลบได้

(4) สามารถค้นหาข้อมูลความคิดเห็นเชิงบวกได้

(5) สามารถค้นหาข้อมูลความคิดเห็นที่เป็นกลางได้

(6) สามารถค้นหาข้อมูลตามชื่อผู้ใช้งานได้

(7) สามารถค้นหาข้อมูลตาม Facebook ID ได้

(8) สามารถส่งข้อมูลความคิดเห็นไปทดสอบประโยคได้

6) ระบบแจ้งเตือนข้อมูลผ่านทาง Line

(1) สามารถแสดง QR Code เพื่อให้ผู้ดูแลระบบสแกนข้อมูลได้

(2) สามารถแจ้งเตือนข้อความกรณีมีความคิดเห็นเชิงลบได้

(3) สามารถส่งลิงค์ต้นทางความคิดเห็นของผู้ใช้งานไปยัง Lineได้

4.3 ขอบเขตด้านเทคโนโลยี พัฒนาระบบโดยใช้

4.3.1 Laravel Framework 5.3 บนภาษาพีเอชพี (PHP) สำหรับพัฒนาระบบ

4.3.2 NodeJS บนภาษา JavaScript สำหรับการทำงานของข้อมูลแบบเรียลไทม์ (Real-time)

4.3.3 MySQL เป็นระบบจัดการฐานข้อมูล

4.3.4 Facebook Graph API สำหรับสืบค้นข้อมูลจากเฟซบุ๊ก

4.3.5 Apache สำหรับเป็นเว็บเซิร์ฟเวอร์

4.3.6 Line Messaging API สำหรับสร้างการแจ้งเตือนด้วยข้อความผ่านไลน์

4.3.7 Digital Ocean สำหรับสร้างเครื่องเซิฟเวอร์บนระบบปฏิบัติการ Debian Jessie

4.3.8 Cloudflare สำหรับทำ Proxy การเชื่อมต่อแบบปลอดภัย

4.4 อุปกรณ์และเครื่องมือที่ใช้ในการดำเนินการ

4.4.1 ฮาร์ดแวร์

1) เครื่องคอมพิวเตอร์ส่วนบุคคลจำนวน 1 เครื่อง ดังรายละเอียดดังนี้
   - Hard disk 256 GB SSD
   - Ram 8 GB
   - CPU Core i5 2.6 GHz
2) โทรศัพท์มือถือจำนวน 1 เครื่อง
3) เครื่องพิมพ์ (Printer) จำนวน 1 เครื่อง

4.4.2 ซอฟต์แวร์

1) Visual Studio Code ใช้ในการเขียนโปรแกรม
2) Source Tree ใช้ในการจัดการข้อมูล Source Code
3) Microsoft Word 2016 สำหรับจัดทำเอกสารรูปเล่ม

5. ประโยชน์ที่คาดว่าจะได้รับ

5.1 ได้ระบบที่สามารถวิเคราะห์ความรู้สึกจากข้อความภาษาไทยแบบเรียลไทม์

5.2 ได้ข้อมูลความรู้สึกของนักศึกษาบนเฟซบุ๊กแบบเรียลไทม์ (Real-time) ที่แม่นยำและถูกต้อง

5.3 ได้ข้อมูลที่แม่นยำและถูกต้องเพื่อนำไปเป็นแนวทางการปรับปรุง พัฒนาหลักสูตร การเรียนการสอน และระบบบริการของมหาวิทยาลัย

6. นิยามศัพท์เฉพาะ

ในการวิจัยการพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ (Real-time) ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ (Naive Bayes Classifier) สำหรับภาษาไทย มีนิยามศัพท์เฉพาะที่สำคัญดังนี้

6.1 การวิเคราะห์ความรู้สึก (Sentiment Analysis) เป็นการวิเคราะห์ความคิดเห็นของบุคคลทั่วไปหรือกลุ่มลูกค้าบนสื่อสังคมออนไลน์ ว่ารู้สึกอย่างไรต่อสินค้าและบริการ โดยรวบรวมข้อความจากสื่อสังคมออนไลน์ แล้วนำมาประมวลผลทางภาษาเพื่อวิเคราะห์และแสดงผลด้วยรูปแบบที่เข้าใจง่าย

6.2 เฟซบุ๊ก (Facebook) เป็นเครื่องมือทางสังคมประเภทหนึ่งที่ทำให้มนุษย์มีการแลกเปลี่ยน แบ่งปันข้อมูลซึ่งกันและกันกับบุคคลอื่นได้อย่างมีประสิทธิภาพ และมีความรวดเร็ว โดยเฟซบุ๊กเปรียบเสมือนชุมชน ๆ หนึ่งในโลกจริง ที่มีการใช้ชื่ออีเมล์ในการเข้าร่วมเป็นสมาชิกเพื่อทำการติดต่อสื่อสารกับคนอื่น ๆ

6.3 ระบบแบบเรียลไทม์ (Real-time System) คือระบบการทำงานแบบทันที ที่มีการตอบสนองทันที เช่น ระบบการรายงานผลแบบเรียลไทม์ (Real-time) คือระบบที่สามารถรายงานผลได้ทันที และตลอดเวลา

6.4 ตัวจำแนกข้อมูลนาอีฟ เบย์ (Naive Bayes Classifier) เป็นอัลกอริทึมในการจำแนกหมวดหมู่โดยใช้หลักการความน่าจะเป็น ที่สามารถคาดการณ์ผลลัพธ์และสามารถอธิบายได้โดยทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็น

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ผู้วิจัยได้ศึกษาค้นคว้าข้อมูล ทฤษฎี และงานวิจัยที่เกี่ยวข้องจากเอกสาร บทความวิชาการและแหล่งสารสนเทศอินเทอร์เน็ต ในหัวข้อดังต่อไปนี้

1. ลักษณะของภาษาไทย
2. การสืบค้นข้อมูลเฟซบุ๊กด้วย Facebook Graph API
3. การตัดคำภาษาไทย (Thai Word Segmentation)
4. การวิเคราะห์ความรู้สึก (Sentiment Analysis)
5. ตัวจำแนกข้อมูลนาอีฟ เบย์ (Naive Bayes Classifier)
6. การแจ้งเตือนผ่านทาง Line ด้วย Line Messaging API
7. ระบบเรียลไทม์ (Real-time System)
8. การทดสอบประสิทธิภาพ (Performance Evaluation)
9. งานวิจัยที่เกี่ยวข้อง

1. ลักษณะของภาษาไทย

ภาษาไทยมีลักษณะที่เขียนคำและพยางค์ติดกันไปทั้งประโยค คำไทยคำหนึ่งจะประกอบไปด้วยพยัญชนะ สระ และวรรณยุกต์ อีกหลายตัวประกอบกัน โดยแต่ละคำและพยางค์จะประกอบกันเข้าเป็นประโยคของภาษาไทย ซึ่งพยัญชนะไทยมีฟที่เป็นพยัญชนะต้นเป็นตัวแรกของคำหรือพยางค์ และพยัญชนะท้ายหรือตัวสะกด โดยพยัญชนะของภาษาไทยมีทั้งหมด 44 รูป ดังตารางที่ 2.1

ตางรางที่ 2.1 พยัญชนะไทย 44 รูป

ก	ศ	ฉ	ญ	ฑ	ต	น	ฝ	ม	ว	ห
ข	ฆ	ช	ฎ	ฒ	ถ	บ	พ	ย	ศ	ฬ
ฃ	ง	ซ	ฏ	ณ	ท	ป	ฟ	ร	ษ	อ
ค	จ	ฌ	ฐ	ด	ธ	ผ	ภ	ล	ส	ฮ

ที่มา : ประเสริฐ ศรีราชพัฒน์ (2555)

สระ ในภาษาไทย มี 21 รูป 32 เสียง ที่ต่างจากสระในภาษาอังกฤษที่มีเพียงแค่ 5 ตัว ได้แก่ a e i o u โดยสระในภาษาไทยมีรูปสระและมีชื่อเรียกต่าง ๆ ดังตารางที่ 2.2

ตารางที่ 2.2 สระ 21 รูป

ที่	รูปสระ	ชื่อสระ	ที่	รูปสระ	ชื่อสระ	ที่	รูปสระ	ชื่อสระ
1	ะ	วิสรรชนีย์	8	็	ฟันหนู	15	อ	ตัว ออ
2	อั	ไม้หันอากาศ	9	อุ	ตีนเหยียด	16	ย	ตัว ยอ
3	อ็	ไม้ไต่คู้	10	อู	ตีนคู้	17	ว	ตัว วอ
4	า	ลากข้าง	11	เ	ไม้หน้า	18	ฤ	ตัว รึ
5	อิ	พินทุ์อิ	12	ใ	ไม้ม้วน	19	ฤๅ	ตัว รือ
6	‘	ฝนทอง	13	ไ	ไม้มลาย	20	ฦ	ตัว ลึ
7	อำ	นิคหิต	14	โ	ไม้โอ	21	ฦๅ	ตัวลือ

ที่มา : ประเสริฐ ศรีราชพัฒน์ (2555)

วรรณยุกต์ คือ เครื่องหมายที่ใช้กำกับคำเพื่อให้มีระดับเสียงต่างกัน ภาษาที่ไม่มีวรรณยุกต์มีแต่สระกับพยัญชนะก็พอแล้วจะอ่านสูง ๆ ต่ำ ๆ อย่างไรก็แล้วแต่ ในภาษาไทยนิยมใช้วรรณยุกต์ด้วย จึงต้องมีอักษรวรรณยุกต์บังคับอีกต่อหนึ่ง ประโยชน์ของวรรณยุกต์ คือช่วยทำให้คำมีความหมายมากขึ้น แตกต่างจากภาษาของชาติอื่น ๆ เช่น ปา ป่า ป้า ป๊า ป๋า (เสียงนี้มี 4 ความหมาย) ต่างจากภาษาอังกฤษไม่ว่าจะออกเสียงสูง - ต่ำ อย่างไร ความหมายยังคงเดิม ตัวอย่างเช่น

“ปา” หมายถึง ขว้างปา โยน

“ป่า” หมายถึง ที่มีต้นไม้ภูเขา และสัตว์

“ป้า” หมายถึง พี่ของพ่อหรือแม่

“ป๊า” หมายถึง พ่อในภาษาบางภาษา

“ป๋า” หมายถึง พ่อในภาษาบางภาษา

ซึ่งวรรณยุกต์ทั้งหมดมี 4 รูป 5 เสียง ดังนี้

1. เสียงสามัญ คือ เสียงกลาง ๆ เช่น กา มา ทา เป็น ชน
2. เสียงเอก ก่า ข่า ป่า ดึก จมูก ตก หมด
3. เสียงโท เช่น ก้า ค่า ลาก พราก กลิ้ง สร้าง

4. เสียงตรี เช่น ก๊า ค้า ม้า ช้าง โน้ต มด

5. เสียงจัตวา เช่น ก๋า ขา หมา หลิว สวย หาม ปิ๋ว จิ๋ว

โดยสรุปแล้ว คำและพยางค์ของภาษาไทยจะประกอบไปด้วย พยัญชนะ + สระ + วรรณยุกต์ เพื่อให้ได้คำที่มีความหมายออกมาให้เข้าใจนั่นเอง

1.1 ชนิดของคำในภาษาไทย คำในภาษาไทยจะแยกเป็นชนิดต่าง ๆ ดังนี้

1.1.1 คำนาม คือ คำที่ใช้เรียกชื่อคน สัตว์ สิ่งของ สถานที่ อาคาร สภาพ และลักษณะทั้งสิ่งมีชีวิต และไม่มีชีวิต ทั้งที่เป็นรูปธรรมและนามธรรม เช่น คน , รถ , หนังสือ, กล้วย เป็นต้น

1.1.2 คำสรรพนาม คือ คำที่ใช้แทนนามในประโยคสื่อสาร โดยใช้คำสรรพนามเพื่อไม่ต้องกล่าวคำนามซ้ำ ๆ เช่น สรรพนามที่ใช้ในการพูด คือ ฉัน ดิฉัน ผม ข้าพเจ้า เรา หนู เป็นต้น

1.1.3 คำกริยา คือ คำที่แสดงอาการ สภาพ หรือการกระทำของคำนาม และคำสรรพนามในประโยค คำกริยาบางคำอาจมี ความหมายสมบูรณ์ในตัวเอง บางคำต้องมีคำอื่นมาประกอบ และบางคำต้องไปประกอบคำอื่นเพื่อขยายความ เช่น ครูยืน น้องนั่งบนเก้าอี้ ฝนตกหนัก เด็ก ๆ หัวเราะ เป็นต้น

1.1.4 คำวิเศษณ์ คือ คำที่ใช้ขยายคำอื่น ได้แก่ คำนาม คำสรรพนาม คำกริยา หรือคำวิเศษณ์ ให้มีความหมายชัดเจนขึ้น เช่น ชนิด ขนาด สี เสียง กลิ่น รส อาการ เป็นต้น

1.1.5 คำบุพบท คือ คำที่เชื่อมคำหรือกลุ่มคำให้สัมพันธ์กันและเมื่อเชื่อมแล้วทำให้ทราบว่า คำ หรือกลุ่มคำที่เชื่อมกันนั้นมีความสัมพันธ์กันอย่างไร ได้แก่ ใน แก่ จน ของ ด้วย โดย เป็นต้น

1.1.6 คำสันธาน คือ คำที่ทำหน้าที่เชื่อมคำกับคำ เชื่อมประโยคกับประโยค เชื่อมข้อความกับข้อความ หรือข้อความให้สละสลวย เช่น คำว่า กับ และ , ทั้ง...และ , ทั้ง...ก็ , ครั้น...ก็ , ครั้น...จึง , พอ...ก็ ตัวอย่างเช่น “พออ่านหนังสือเสร็จก็เข้านอน”

1.1.7 คำอุทาน คือ คำที่เปล่งออกมาเพื่อแสดงอารมณ์หรือความรู้สึกของผู้พูด มักจะเป็นคำที่ไม่มีความหมาย แต่เน้นความรู้สึกและอารมณ์ของผู้พูด คำอุทานนี้ แบ่งเป็น 3 ลักษณะ

1.1.7.1 เป็นคำ เช่น โอ๊ย ว้าย แหม โถ เป็นต้น

1.1.7.2 เป็นวลี เช่น พุทโธ่เอ๋ย คุณพระช่วย ตายล่ะว้า

1.1.7.3 เป็นประโยค เช่น ไฟไหม้เจ้าข้า เฮ้ยป้าถูกรถชน

1.2 ประโยคภาษาไทย

ในประโยคหนึ่ง ๆ จะต้องมีภาคประธานและภาคแสดงเป็นหลัก และอาจมีคำขยายส่วนต่าง ๆ ได้ ซึ่งส่วนประกอบของประโยค จะประกอบด้วย ภาคประธาน ที่ทำหน้าที่เป็นผู้กระทำ และภาคแสดง ที่เป็นกลุ่มคำที่ประกอบไปด้วยบทกริยา บทกรรมและส่วนเติมเต็ม บทกริยาทำหน้าที่เป็นตัวกระทำหรือตัวแสดงของประธาน ส่วนบทกรรมทำหน้าที่เป็นผู้ถูกกระทำ และส่วนเติมเต็มทำหน้าที่เสริมใจความของประโยคให้สมบูรณ์ คือทำหน้าที่คล้ายบทกรรม แต่ไม่ใช้กรรม เพราะมิได้ถูกกระทำ นอกจากนี้ประโยคในภาษาไทยยังแบ่งเป็น 3 ชนิด ตามโครงสร้างการสื่อสารดังนี้

1.2.1 ประโยคความเดียว คือ ประโยคที่มีข้อความหรือใจความเดียว เป็นประโยคที่มีภาคประโยคเพียงบทเดียว และมีภาคแสดงหรือกริยาสำคัญเพียงบทเดียว หากภาคประธานและภาคแสดงเพิ่มบทขยายเข้าไป ประโยคความเดียวนั้นก็จะเป็นประโยคความเดียวที่ซับซ้อนยิ่งขึ้น

1.2.2 ประโยคความรวม คือ ประโยคที่รวมเอาโครงสร้างประโยคความเดียวตั้งแต่ 2 ประโยคขึ้นไปเข้าไว้ในประโยคเดียวกัน โดยมีคำเชื่อมหรือคำสันธานทำหน้าที่เชื่อมประโยคเหล่านั้นเข้าด้วยกัน ซึ่งประโยคความรวมยังแบ่งออกเป็นอีก 4 ประเภท ดังนี้

1.2.2.1 ประโยคที่มีความคล้อยตามกัน ประโยคความรวมชนิดนี้ประกอบด้วยประโยคเล็กตั้งแต่ 2 ประโยคขึ้นไป มีเนื้อความคล้อยตามกันในแง่ของความเป็นอยู่ เวลา และการกระทำ

ตัวอย่าง เช่น “ขอแสดงความยินดีกับบัณฑิตทุกท่านค่ะ เป็นความภาคภูมิใจค่ะที่จบจากที่นี่ มสธ.”

1.2.2.2 ประโยคที่มีความขัดแย้งกัน ประโยคความรวมชนิดนี้ ประกอบด้วยประโยคเล็ก 2 ประโยค มีเนื้อความที่แย้งกันหรือแตกต่างกันในการกระทำ หรือผลที่เกิดขึ้น

ตัวอย่าง เช่น “ได้ใบรับรองจบมาแล้วแต่หมดอายุ”

1.2.2.3 ประโยคที่มีความให้เลือก ประโยคความรวมชนิดนี้ ประกอบด้วยประโยคเล็ก 2 ประโยคและกำหนดให้เลือกอย่างใดอย่างหนึ่ง

ตัวอย่าง เช่น “เรียนจบได้วุฒิอะไร หรือแค่ประกาศนียบัตร ?”

1.2.2.4 ประโยคที่มีความเป็นเหตุเป็นผลแก่กัน ประโยคความรวมชนิดนี้ประกอบด้วยประโยคเล็ก 2 ประโยค ประโยคแรกเป็นเหตุประโยคหลังเป็นผล

ตัวอย่าง เช่น “มาเป็นพี่น้องกันนะคะ มีรุ่นพี่ที่น่ารักคอยแนะนำน้อง ๆ เราจะจบไปพร้อมกัน”

1.2.3 ประโยคความซ้อน คือ ประโยคที่มีใจความสำคัญเพียงใจความเดียว ประกอบด้วยประโยคความเดียวที่มีใจความสำคัญ และมีประโยคความเดียวที่มีใจความเป็นส่วนขยายส่วนใดส่วนหนึ่งของประโยคหลัก เป็นประโยคย่อยซ้อนอยู่ในประโยคหลัก โดยทำหน้าที่แต่งหรือประกอบประโยคหลัก เช่น ครูดุนักเรียนไม่ทำการบ้าน ครูดุนักเรียน เป็นประโยคหลัก นักเรียนไม่ทำการบ้าน เป็นประโยคย่อยทำหน้าที่เป็นบทกรรม

ประโยคต่าง ๆ ที่ใช้ในการสื่อสารย่อมแสดงถึงเจตนาของผู้ส่งสาร เช่น บอกกล่าว เสนอแนะ อธิบาย ซักถาม ขอร้อง วิงวอน สั่งห้าม เป็นต้น หากจะแบ่งประโยคตามหน้าที่หรือลักษณะที่ใช้ในการสื่อสาร จะเห็นได้ว่าภาษาไทยนั้นมีลักษณะเฉพาะทางโครงสร้างที่ซับซ้อน เมื่อเทียบกับภาษาอังกฤษหรือภาษาจีนเมื่อมีการนำคอมพิวเตอร์เข้ามาประมวลผล โดยประโยคภาษาไทยมีลักษณะเฉพาะของภาษา ดังนี้

1) ไม่มีสัญลักษณ์พิเศษเพื่อบอกขอบเขตของคำ
เช่น “หญิงสาวนั่งตากลม” “หลวงตามหาบัว”

2) หนึ่งหน่วยคำประกอบด้วยตัวอักษรมากกว่า 1 ตัว จึงทำให้เกิดความคลุมเครือในการแบ่งขอบเขตของคำได้มากขึ้น
เช่น “มาก่อว่า” “หาก่อว่า” “ขนบนอก” “ดูดวง”

3) ไม่มีชุดอักษรพิเศษที่บอกถึงคำยืม คำทับศัพท์ หรือ ชื่อเฉพาะ
เช่น “หมูกินข้าว” “จอห์น สัน ซื้อ แป้ง เด็ก แคร์” “ขอแชร์นะครับ”

2. การสืบค้นข้อมูลเฟซบุ๊กด้วย Facebook Graph API

เฟซบุ๊ก (Facebook) (วิกิพีเดีย สารานุกรมเสรี, 2561) เป็นบริการเครือข่ายสังคมสัญชาติอเมริกัน สำนักงานใหญ่อยู่ที่ เมนโลพาร์ก รัฐแคลิฟอร์เนีย เฟซบุ๊กก่อตั้งเมื่อวันพุธที่ 4 กุมภาพันธ์ ค.ศ. 2004 โดยมาร์ก ซักเคอร์เบิร์ก และเพื่อนร่วมห้องภายในมหาวิทยาลัย และเหล่าเพื่อนในมหาวิทยาลัยฮาร์วาร์ด พร้อมโดยสมาชิกเพื่อนผู้ก่อตั้ง Eduardo Saverin, Andrew McCollum, Dustin Moskovitz และ Chris Hughes ในท้ายที่สุดเว็บไซต์มีการเข้าชมอย่างจำกัด ทำให้ใช้เพียงแค่ดกลุ่มนักศึกษาภายในมหาวิทยาลัยฮาร์วาร์ด แต่ภายหลังได้ขยายเพิ่มจำนวนมหาวิทยาลัยในพื้นที่บอสตัน ไอวีลีก และมหาวิทยาลัยสแตนฟอร์ด แล้วค่อย ๆ รับรองมหาวิทยาลัยอื่น ๆ และต่อมาก็รับรองโรงเรียนมัธยมศึกษา โดยเฟซบุ๊กให้อนุญาตเยาวชนอายุ 13 ปีขึ้นไป ทั่ว

โลกสามารถสมัครสมาชิกได้ภายในเว็บไซต์ โดยไม่ต้องมีเอกสารหลักฐานใด ๆในการสมัคร

Facebook Graph API คือ วิธีการในการรับส่งข้อมูลระหว่าง Facebook API กับไคลเอนต์ ซึ่งทำงานอยู่บนโปรโตคอล (Protocal) HTTP ไคลเอนต์ (Client) สามารถใช้งานการสืบค้นข้อมูลความคิดเห็นบนเฟซบุ๊ก ข้อมูลส่วนตัวของผู้ใช้งาน ข้อมูลความคิดเห็นที่แสดงบนเฟซบุ๊กเพจ ข้อความที่ส่งผ่านทาง Massager จัดการโฆษณา อัพโหลดรูปภาพ และทำงานอื่น ๆ ผ่านทาง Facebook Grap API การเรียกใช้ข้อมูลด้วย Facebook Grap API มีเงื่อนไขที่เฟซบุ๊กได้ออกแบบเพื่อป้องกัน นักพัฒนาที่ไม่ได้รับอนุญาตในการใช้งานข้อมูล ทั้งนี้ก็เพื่อคัดกรองบุคคลหรือหน่วยงานที่จะเรียกใช้ข้อมูล การเข้าถึงข้อมูลการแสดงความคิดเห็นบนเฟซบุ๊กเพจนั้น สิ่งที่นักพัฒนาระบบจำต้องปฏิบัติมีขั้นตอนเบื้องต้นดังนี้

1) ผู้ใช้งานเฟซบุ๊กต้องเป็นนักพัฒนาโดยใช้งานในส่วนของนักพัฒนา
2) ต้องมีเฟซบุ๊กแอปพลิเคชัน (Facebook Aplication) บนเฟซบุ๊ก
3) ต้องได้รับอนุญาตในการใช้งานข้อมูลของเฟซบุ๊กเพจ

Access Token คือกุญแจในการร้องขอใช้งานข้อมูลระหว่าง Client กับ Server บนมาตรฐาน Oauth (Open Authentication) คือ เป็นมาตรฐานที่ Application จะใช้ติดต่อกับเครื่อง Client ในแบบการเข้าระบบผ่านตัวแทนที่ปลอดภัย ที่ใช้สำหรับการกำหนดสิทธิ์ให้ Application หนึ่งสามารถร้องขอทรัพยากรของผู้ใช้จาก Application หนึ่งได้โดยที่ Application นั้นไม่จำเป็นต้องทราบรหัสผ่านของผู้ใช้ Application ที่ร้องขอทรัพยากรจะได้รับสิ่งที่เรียกว่า Access Token กลับมาหลังจากเข้าระบบสำเร็จ เพื่อใช้แทน Username และ Password เพื่อนำไปใช้กับบริการอื่น ๆ ทำให้มีความปลอดภัยมากขึ้น รวมถึงบอกว่ามีสิทธิ์ทำอะไรได้บ้างกับบริการนั้น ๆ โดยแนะนำให้ใช้คู่กับโปรโตคอล (Protocol) Https อีกชั้นเพื่อความปลอดภัยสูงสุด สำหรับ Token ประเภทต่าง ๆ ที่ใช้งานในเฟซบุ๊กมีดังตารางที่ 2.3

ตารางที่ 2.3 ประเภทของ Access Token Facebook

ประเภทของโทเค็นการเข้าถึง	คำอธิบาย
User Access Token	โทเค็นของผู้ใช้เป็นประเภทที่ใช้กันมากที่สุด โทเค็นการเข้าถึงประเภทนี้จำเป็นต้องมีเมื่อแอปพลิเคชันเรียก API เพื่ออ่าน แก้ไข หรือเขียนข้อมูลบางอย่างของผู้ใช้บน Facebook แทนผู้ใช้รายนั้น โทเค็นการเข้าถึงของผู้ใช้นั้นมักจะได้รับผ่านข้อความโต้ตอบเพื่อเข้าสู่ระบบ และกำหนดให้ผู้ใช้ต้องอนุญาตให้แอปพลิเคชันของผู้พัฒนา รับโทเค็นนั้น
App Access Token	จำเป็นต้องมีโทเค็นการเข้าถึงประเภทนี้เพื่อแก้ไขและอ่านการตั้งค่าแอปพลิเคชัน และยังสามารถใช้เพื่อเผยแพร่การดำเนินการ Open Graph ได้ด้วย โทเค็นประเภทนี้สร้างขึ้นโดยใช้ข้อมูลลับที่แอปพลิเคชันและ Facebook ตกลงไว้แล้ว
Page Access Token	โทเค็นการเข้าถึงประเภทนี้คล้ายกับโทเค็นการเข้าถึงของผู้ใช้ แต่ต่างกันที่โทเค็นการเข้าถึงของเพจให้สิทธิ์การอนุญาตไปยัง API ที่อ่าน เขียน หรือแก้ไขข้อมูลของเพจ Facebook ได้ หากต้องการโทเค็นการเข้าถึงของเพจ นักพัฒนาจะต้องเริ่มโดยรับโทเค็นการเข้าถึงของผู้ใช้และขอสิทธิ์การอนุญาต manage_pages หลังจากมีโทเค็นการเข้าถึงของผู้ใช้แล้ว นักพัฒนาจะได้รับโทเค็นการเข้าถึงของเพจผ่าน API กราฟ
Client Token	โทเค็นของไคลเอ็นต์เป็นตัวระบุที่นักพัฒนาสามารถฝังลงในไบนารีรูปแบบเดิมบนมือถือหรือแอปพลิเคชันบนเดสก์ท็อปเพื่อระบุแอปพลิเคชันของนักพัฒนาได้ โทเค็นของไคลเอ็นต์นั้นไม่ควรใช้เป็นตัวระบุข้อมูลลับ เนื่องจากโทเค็นนี้ถูกฝังไว้ในแอปพลิเคชัน โทเค็นของไคลเอ็นต์นั้นใช้เพื่อเข้าถึง **API** ระดับแอปพลิเคชัน

Page Access Token คือ โทเค็นการเข้าถึงของเฟซบุ๊กเพจนั้นใช้ในการเรียก API กราฟ เพื่อจัดการเฟซบุ๊กเพจ ผู้ดูแลเฟซบุ๊กเพจจะต้องให้สิทธิ์การอนุญาตแบบขยาย ที่เรียกว่า manage_pages เพื่อสร้างโทเค็นการเข้าถึงของเฟซบุ๊กเพจ หลังจากได้รับสิทธิ์การอนุญาตนี้แล้ว นักพัฒนาสามารถเรียกดูโทเค็นการเข้าถึงของเฟซบุ๊กเพจโดยใช้โทเค็นการเข้าถึงของผู้ใช้ร่วมกับสิทธิ์การอนุญาตที่กำหนดให้ และส่งคำขอใช้ข้อมูลผ่านทาง Facebook Graph API ซึ่งจะส่งคืนรายการเฟซบุ๊กเพจที่ผู้ใช้ดูแลพร้อมข้อมูลอื่น ๆ เกี่ยวกับเฟซบุ๊กเพจ เช่น หมวดหมู่เพจ สิทธิ์การอนุญาตที่ผู้ดูแลมีในเพจนั้น และโทเค็นการเข้าถึงของเพจ ตัวอย่างข้อมูล JSON เฟซบุ๊กเพจดังภาพที่ 2.1


```
{
 "data": [
  {
   "category": "Product/service",
   "name": "Sample Page",
   "access_token": "{access-token}",
   "id": "1234567890",
   "perms": [
    "ADMINISTER",
    "EDIT_PROFILE",
    "CREATE_CONTENT",
    "MODERATE_CONTENT",
    "CREATE_ADS",
    "BASIC_ADMIN"
   ]
  },
}
```


ภาพที่ 2.1 ตัวอย่างข้อมูล JSON ของเฟซบุ๊กเพจ

นักพัฒนาสามารถใช้โทเค็นการเข้าถึงเฟซบุ๊กเพจในการเรียก API ในนามของเพจ เช่น นักพัฒนาสามารถโพสต์การอัพเดตสถานะในเฟซบุ๊กเพจ แทนที่จะโพสต์ในไทม์ไลน์ของผู้ใช้ หรืออ่านข้อมูลเชิงลึกของเฟซบุ๊กเพจโทเค็นการเข้าถึงของเฟซบุ๊กเพจจะไม่ซ้ำกันในแต่ละเฟซบุ๊กเพจ ผู้ดูแลระบบแต่ละคน และแอปพลิเคชันแต่ละแอปพลิเคชัน

3. การตัดคำภาษาไทย (Thai Word Segmentation)

การตัดคำภาษาไทย (Thai Words Segmentation) (กานดา รุณนะพงศา, 2549) เป็นกระบวนการพื้นฐานของการประมวลผลภาษาธรรมชาติ ในการตัดคำภาษาไทยนั้นได้มีผู้คิดค้นวิธีที่จะแยกคำแต่ละคำออกจากประโยคซึ่งมีการเขียนติดกันไปอย่างต่อเนื่องทั้งประโยค ซึ่งจะแตกต่างจากประโยคในภาษาอังกฤษที่ได้มีการแยกคำออกจากประโยคเป็นคำ ๆ ทำให้สะดวก เนื่องจากสามารถตรวจสอบขอบเขตของคำในประโยคได้ง่าย โดยมีการบอกขอบเขตของคำเป็นการเว้นวรรค และจุดทศนิยมในการบอกการสิ้นสุดของประโยค แต่วิธีการนี้ไม่สามารถนำมาใช้ได้กับคำและประโยคในภาษาไทย เนื่องจากภาษาไทยไม่มีสัญลักษณ์หรือเครื่องหมายใด ๆ ที่เป็นตัวบอกขอบเขตของคำหรือประโยค ดังนั้นจึงเป็นการยากที่จะกำหนดตำแหน่งของคำแต่ละคำได้ จึงได้มีเทคนิคและวิธีการที่นำไปใช้ในการตัดคำในภาษาไทย โดยได้รับการพัฒนาจากหน่วยงานวิจัยต่าง ๆ ทั้งของภาครัฐและภาคเอกชน ได้มีการพัฒนาแนวคิดและวิธีการต่าง ๆ เพื่อนำมาใช้ในการตัดคำในภาษาไทยให้ถูกต้องมากที่สุด โดยแต่ละวิธีการต่างก็ให้ผลในด้านของความถูกต้อง ความรวดเร็วของการทำงาน และปริมาณการใช้ทรัพยากรของระบบ แตกต่างกันไป ซึ่งเทคนิคและวิธีการตัดคำภาษาไทยที่ใช้กันอยู่ทั่ว ๆ ไปสามารถแบ่งได้เป็นหลายวิธี ดังนี้

3.1 หลักการตัดคำโดยใช้กฎ (Rule Based Approach)

การตัดคำโดยใช้กฎ เป็นวิธีการตัดคำแบบแรกๆ ที่ถูกคิดขึ้นมาเพื่อใช้ตัดคำในภาษาไทย โดยเป็นการตรวจสอบกฎเกณฑ์ทางอักขระวิธีที่กำหนดลักษณะการประสมอักษร ลักษณะการเว้นวรรค และการขึ้นย่อหน้า เพื่อใช้เป็นเกณฑ์ในการระบุขอบเขตของคำ ตัวอย่างเช่น

1) การขึ้นย่อหน้าใหม่เป็นตัวบอกถึงการสิ้นสุดข้อความ หรือประโยค

2) การเว้นวรรคเป็นตัวบอกถึง ความเป็นไปได้ที่จะเป็นการสิ้นสุดคำ หรือสิ้นสุดประโยค

3) กฎทางอักขระวิธีนี้เป็นตัวชี้ให้เห็นถึงความเป็นไปได้ของการตัดคำในตำแหน่งนั้น ๆ

โดยวิธีการการตัดคำโดยใช้กฎนี้จะมีข้อจำกัดในการทำงาน คือ ผลของการตัดคำอาจได้เป็นกลุ่มคำที่สามารถตัดคำแยกย่อยออกไปได้อีก ทำให้ความถูกต้องของการตัดคำค่อนข้างต่ำ และยังไม่สามารถแก้ปัญหาความกำกวมของพยัญชนะต้นและตัวสะกดได้ ตัวอย่างเช่น คำว่า “ตากลม” ที่ยังมีปัญหาของตัว “ก” แต่ข้อดีของวิธีการนี้คือมีความรวดเร็วในการทำงานและใช้ทรัพยากรน้อย จึงเหมาะกับการใช้ได้กับงานแค่บางประเภทเท่านั้น

3.2 หลักการตัดคำโดยใช้พจนานุกรม (Dictionary Approach)

การตัดคำโดยใช้พจนานุกรม เป็นวิธีที่มีการพัฒนาต่อมาจากการใช้กฎในการตัดคำ โดยวิธีการนี้จะทำการเก็บคำภาษาไทยไว้ในพจนานุกรมโดยใช้ผู้เชี่ยวชาญด้านภาษาในการดำเนินการและตรวจสอบคำให้ถูกต้องก่อนจัดเก็บลงพจนานุกรม เพื่อไม่ให้เกิดข้อผิดพลาดในการทำงานของระบบ แล้วจึงนำข้อความที่รับเข้ามา ไปค้นหาและเปรียบเทียบสายอักขระกับคำที่เก็บไว้ในพจนานุกรม เพื่อหาว่าข้อความดังกล่าวควรตัดคำในบริเวณใด และประกอบไปด้วยคำใดบ้าง วิธีนี้จะทำให้ได้ความถูกต้องในการตัดคำสูงกว่าการใช้กฎ แต่จะใช้เวลาในการทำงานมากกว่า ซึ่งความเร็วในการทำงานจะขึ้นอยู่กับจำนวนคำ ที่มีอยู่ในพจนานุกรมด้วย

รัทยานนท์ สมปรารถนา (2535)ได้เสนอหลักการทำงานทางเทคนิคคือ ระบบจะค้นหาคำจากข้อความที่ป้อนเข้าไป เทียบกับคำในพจนานุกรม เพื่อนำคำแต่ละคำไปจัดเก็บไว้ในแถวลำดับหรืออาเรย์ (Array) โดยการค้นหาจะเริ่มค้นหาจากข้อความแรก แล้วนำคำแรกที่เทียบเจอในพจนานุกรมไปจัดเก็บเอาไว้ในช่องที่หนึ่ง แล้วจึงตัดคำที่พบออกไปจากข้อความ จากนั้นนำข้อความที่เหลือจากการตัดคำออกไป มาเทียบคำกับพจนานุกรมอีกครั้ง เพื่อนำแต่ละคำที่เทียบเจอไปจัดเก็บไว้ในอาเรย์ช่องถัดไป ทำอย่างนี้ไปเรื่อยๆ จนหมดข้อความ (ถ้าให้อาเรย์ช่องสุดท้ายเป็นช่องที่ n) จากนั้นจึงเปรียบเทียบย้อนกลับโดยเทียบคำที่อยู่ในอาเรย์ แต่ละช่องกับคำในพจนานุกรมตั้งแต่ช่องที่ n, n-1, n-2, …ไปเรื่อย ๆ จนถึงอาเรย์ที่อยู่ในช่องแรก เพื่อค้นหาว่าในแต่ละช่องของอาเรย์สามารถตัดคำเป็นคำอื่นที่ต่างออกไปจบการทำงาน ดังนั้นผลที่ได้ออกมาจากการตัดคำจะเป็นรูปแบบของความเป็นไปได้ของการตัดคำทั้งหมดจากข้อความที่ป้อนเข้าไปได้หรือไม่ หากพบว่าในอาเรย์ช่องใด (สมมติให้เป็นช่องที่ i โดยที่ i <= n) สามารถตัดคำในรูปแบบอื่นได้อีก ก็จะตัดคำให้เป็นในรูปแบบใหม่นั้น แล้วจึงเทียบคำต่อไปในอาเรย์ช่องถัดไป (ช่องที่ i+1) จนหมดข้อความ การทำงานดังกล่าวจะทำย้อนกลับไปจนถึงอาเรย์ช่องแรกสุด

ปัญหาที่พบในวิธีการตัดคำโดยใช้พจนานุกรม คือ เป็นการยากที่จะเก็บคำทุกคำในภาษาไทยลงไปในพจนานุกรมได้ทั้งหมดโดยเฉพาะคำวิสามานยนามหรือคำเฉพาะ เช่น ชื่อคน ชื่อ

สถานที่ ตัวเลข วันที่หรือคำที่เกิดขึ้นใหม่ นอกจากนั้นแล้ว วิธีการนี้จะทำให้สิ้นเปลืองทรัพยากรพื้นที่ของหน่วยความจำหลักค่อนข้างมาก เนื่องจากคำในพจนานุกรมทั้งหมด และอาเรย์ของคำที่ตัดมาได้จะถูกเก็บเอาไว้ในหน่วยความจำหลักทั้งหมด

หลักการตัดคำโดยใช้พจนานุกรมนี้สามารถตัดคำภาษาไทยได้ถูกต้องมากกว่าการใช้กฎเพราะฉะนั้นวิธีการตัดคำแบบนี้จึงได้รับความนิยมมากกว่า และมีการพัฒนาวิธีการที่ใช้ในการตัดคำอื่น ๆ โดยใช้หลักการของพจนานุกรมช่วยอีกด้วย เช่น

3.2.1 การเทียบคำที่ยาวที่สุด (Longest Matching) วิธีการเทียบคำที่ยาวที่สุด เป็นการตรวจสอบสายอักขระที่นำเข้ามาจากซ้ายไปขวา จากนั้นนำไปเปรียบเทียบกับคำที่มีอยู่ในพจนานุกรม จึงต้องใช้พจนานุกรมช่วยให้รู้จักคำภาษาไทย ว่าสายอักขระดังกล่าวเป็นหนึ่งคำในพจนานุกรมหรือไม่ หากไม่พบ ก็จะทำต่อไปเรื่อย ๆ จนจบข้อความ จนกว่าสายอักขระที่ตรวจสอบจะสามารถเทียบเป็นคำในพจนานุกรมได้ ก็จะทำเครื่องหมายเพื่อเป็นจุดย้อนกลับจากนั้นเริ่มทำงานจากจุดย้อนกลับนั้นเพื่อตรวจสอบสายอักขระที่เหลือว่าสามารถตัดสายอักขระใดต่อไป เพื่อให้เป็นคำได้หรือไม่หากตัวเลือกในตอนแรกนี้สามารถทำให้ขั้นตอนวิธี (Algorithm) ค้นหาคำที่เหลือได้ตัวเลือกนี้ก็จะเป็นคำแรกของข้อความได้จริง ไม่เช่นนั้นขั้นตอนวิธีก็จะกลับไปยังจุดย้อนกลับที่ทำเครื่องหมายไว้ เพื่อแก้ไขคำแรกใหม่จากนั้นก็จะเริ่มทำงานต่อไป โดยเริ่มจากจุดย้อนกลับ หากยังไม่สามารถเทียบสายอักขระให้ตรงกับพจนานุกรมได้ก็จะลดตัวอักษรลงทีละตัวจนกว่าจะเทียบคำในพจนานุกรมให้ตรงกันได้ และทำงานในรูปแบบนี้ต่อไปจนจบข้อความ ดังต่อไปนี้

ตารางที่ 2.4 ตัวอย่างการแบ่งคำด้วยวิธีการเทียบคำที่ยาวที่สุด (Longest Matching)

คำ	**คำที่แบ่งได้**	**คำที่ถูกเลือก**
รองเท้า	1) ร | องเท้า	รอง | เท้า
	2) รอ | งเท้า	
	3) รอง | เท้า	
	4) รองเ | ท้า	

ขั้นตอนที่ 1 แยกอักขระทุกตัวออกจากกัน จากประโยค “รองเท้า” เมื่อแยกออกมาจะได้คำอักขระดังนี้

ร	อ	ง	เ	ท	้	า

ขั้นตอนที่ 2 เทียบอักขระกับฐานข้อมูลคลังคำศัพท์ (Corpus) จากประโยค “รองเท้า” สามารถเทียบคำได้ทั้งหมด 3 คำด้วยกัน คือคำว่า “รอ” , “รอง” และ “รองเท้า”

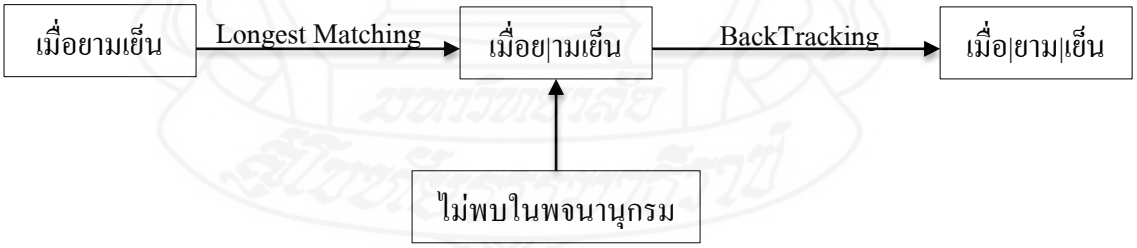
ขั้นตอนที่ 3 เลือกคำที่มีความยาวมากที่สุด (Longest Matching) โดยคำว่า “รองเท้า”

เป็นคำที่ยาวที่สุด จากนั้นคำว่า “รองเท้า” จะถูกตัดออกจากประโยคเพื่อหาคำต่อไปในประโยคด้วยวิธีการแบบเดิมซ้ำ ๆ จนประโยคไม่มีข้อความให้แยกได้อีก

วิธีการนี้ให้ความถูกต้องของการตัดคำได้ค่อนข้างดี แต่ยังมีข้อด้อยที่เกิดจากลักษณะความพยายามที่จะตัดคำให้ได้ยาวที่สุดของอัลกอริทึมแบบละโมบ (Greedy Matching) คือ การเลือกคำที่ยาวเกินไปตั้งแต่ต้น มีผลทำให้การตัดคำที่ตามมาผิดเพี้ยนไปด้วย ตัวอย่างของข้อความที่ป้อนเข้าไปเช่น “ไปหามเหสี” จะตัดคำได้เป็น ไป | หาม | เห | สี เสมอ เนื่องจากอัลกอริทึมจะพบคำว่า “หาม” ก่อนคำว่า “หา” ทำให้การตัดคำสำหรับคำต่อ ๆ ไปผิดไปด้วย

3.2.2 การตัดคำให้ได้จำนวนคำและคำที่ไม่มีในพจนานุกรมน้อยที่สุด (Maximal - Matching)

วิธีการตัดคำให้ได้จำนวนคำและคำที่ไม่มีในพจนานุกรม วิธีจากงานวิจัยของ วิรัชศร เลิศล้ำ วาณิช (2536) เพื่อเป็นการแก้ปัญหาที่เกิดขึ้นในวิธีการเทียบคำที่ยาวที่สุด วิธีการนี้ได้พัฒนาจากขั้นตอนและวิธีของการเทียบคำที่ยาวที่สุด เริ่มจากการหาทางเลือกของรูปแบบการตัดคำทั้งหมดที่เป็นไปได้ก่อน โดยย้อนกลับ (Backtracking) ทีละคำหลังจากได้คำตอบจากวิธีการเทียบคำที่ยาวที่สุด แล้วจึงเลือกทางเลือกที่มีจำนวนคำน้อยที่สุด ดังตัวอย่างภาพที่ 2.2



ภาพที่ 2.2 ตัวอย่างการแบ่งคำด้วยวิธีการย้อนกลับ (Backtracking)

นอกจากนี้ Surapant Meknavin (2000) ได้กล่าวไว้ว่า การที่จะค้นหาทุกทางเลือกที่เป็นไปได้นั้น อาจทำให้ต้องเสียเวลาในการคำนวณที่ค่อนข้างมาก แต่ก็มีวิธีที่สามารถลดเวลาลงได้โดยโปรแกรมแบบพลวัติ (Dynamic Programming) ตัวอย่าง เช่นหากป้อนข้อความ “ไปหามเหสี” เข้าไป ขั้นตอนวิธีนี้จะหาทางเลือกทั้งหมดของรูปแบบการตัดคำที่เป็นไปได้ ซึ่งได้แก่

ไป | หาม | เห | สี

ไป | หา [ม] | เห | สี

ไป | หา | มเหสี

โดยในขั้นแรก จะใช้ขั้นตอนวิธีของการเทียบคำที่ยาวที่สุด และจะได้คำตอบของการตัดคำเป็น “ไป | หาม | เห | สี” ก่อน หลังจากนั้นจึงเริ่มย้อนกลับโดยเริ่มจากคำแรก พบว่า คำที่สอง “หาม” สามารถแบ่งได้เป็น “หา | ม” ได้โดยเมื่อแบ่งเป็น “หา” แล้ว ทำให้สายอักขระที่เหลือ “มเหสี” สามารถตัดคำได้อีก 2 แบบ คือ “มเหสี” และ “[ม] | เห | สี” วิธีการนี้จะให้เลือกข้อความที่แบ่งแล้วมีจำนวนคำน้อยที่สุด คือ แบบที่ 2 ซึ่งมีจำนวนคำที่ตัดได้ 3 คำ ในขณะที่แบบที่ 1 มี 4 คำ ส่วนในกรณีที่มีจำนวนคำที่เท่ากันก็จะใช้วิธีการตัดคำแบบยาวที่สุด (Longest Matching) เข้ามาช่วย เช่น ข้อความว่า "เธอนั่งตากลมที่ระเบียง" ซึ่งจะสามารถแบ่งคำได้ 2 แบบ คือ

เธอ | นั่ง | ตาก | ลม | ที่ | ริม | ระเบียง

เธอ | นั่ง | ตา | กลม | ที่ | ริม | ระเบียง

ทั้ง 2 แบบมีจำนวนคำที่เท่ากันจึงเลือกแบบที่ 1 โดยเปรียบเทียบจากคำที่ต่างกันที่ตัดได้นั้น (ตา / ตาก) จะเห็นได้ว่า ตากมีตัวอักษรมากกว่า โดยวิธีการนี้จะได้ประสิทธิภาพในการแบ่งคำมากกว่าการแบ่งคำด้วยวิธีการเทียบคำที่ยาวที่สุด (Longest Matching) เพราะสามารถแก้ปัญหาในเรื่องของข้อจำกัดในการพยายามเลือกคำให้ยาวที่สุดตั้งแต่ต้น เพราะจะพิจารณาทางเลือกของการตัดคำที่เป็นไปได้ทั้งหมดก่อนที่จะเลือกการตัดคำ

สำหรับการตัดคำภาษาไทยในปัจจุบันนิยมใช้พจนานุกรม LEXiTRON Data 2.0 ซึ่งเป็นพจนานุกรมอิเล็กทรอนิกส์ที่ถูกพัฒนาโดยนักวิจัยจากศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติหรือเนคเทค (NECTEC) เป็นการนำเทคโนโลยีฐานข้อมูลขนาดใหญ่เข้ามาช่วยในการวิจัยและพัฒนาในสาขาการประมวลผลภาษาธรรมชาติ เป็นการใช้เทคโนโลยีทางคอมพิวเตอร์รวบรวมและคัดเลือกคำ ประโยค หรือข้อความ ที่ใช้จริงและมีอัตราการปรากฏสูงในบริบทต่าง ๆ ของการใช้ภาษา จากแหล่งข้อมูลและข่าวสารที่เผยแพร่ทางอินเทอร์เน็ตและแหล่งข้อมูลอื่น ๆ ที่เชื่อถือได้ เช่น วรรณกรรม บทความ เอกสารทางวิชาการ ข้อมูลข่าวสารจากหนังสือพิมพ์ เป็นต้น

3.3 การตัดคำโดยใช้คลังข้อมูล (Corpus Based Approach)

คลังข้อมูลทางภาษา (Corpus) เป็นการเก็บรวมรวมคลังคำศัพท์ไว้ในฐานข้อมูลอิเล็กทรอนิกส์เพื่อใช้ในการเทียบคำตามหลักภาษา โดยส่วนใหญ่จะถูกเลือกใช้เกี่ยวกับงานด้านการประมวลผลทางภาษา และหลักการตัดคำโดยใช้คลังข้อมูลเป็นการนำวิธีการทางสถิติ (Statistical

Techniques) เข้ามาใช้ในการประมวลผลภาษาธรรมชาติ โดยการใช้คลังข้อมูลทางภาษา (Corpus) ที่ใช้แรงงานคนในการเตรียมข้อมูลมาเป็นฐานความรู้สำหรับเก็บค่าความถี่ที่ใช้ในการตัดคำ

3.3.1 การตัดคำแบบคำนวณเชิงสถิติเพื่อหาความเป็นไปได้ (Probabilistic Model) วิธีการนี้เป็นการนำเอาค่าสถิติการเกิดของคำและลำดับของหน้าที่ของคำ (Part of Speech) เข้ามาช่วยในการคำนวณหาความน่าจะเป็นเพื่อที่จะใช้เลือกแบบที่มีโอกาสเกิดมากที่สุด วิธีการนี้สามารถจะตัดคำได้ดี แต่ข้อจำกัดของวิธีการนี้คือจะต้องมีคลังข้อมูลทางภาษา (Corpus) ที่มีการตัดคำที่ถูกต้องและกำหนดหน้าที่ของคำให้เพื่อที่จะได้นำไปใช้ในการสร้างสถิติ

3.3.2 การตัดคำแบบคุณลักษณะ (Feature - Based Approach) วิธีการนี้จะพิจารณาจากบริบทของคำ (Context Words) หรือคำแวดล้อมและการเกิดร่วมกันของคำหรือหน้าที่ของคำ (Collocation) เข้ามาช่วยในการตัดคำ ตัวอย่างเช่น

“ตากลม” ถ้าพบคำว่า “โต” ในบริบทก็จะสามารถตัดคำได้ว่า ตา | กลม

“มากว่า” ถ้าในบริบทที่ตามมาเป็น “ตัวเลข” ก็สามารถตัดคำได้ว่า มาก | ว่า

วิธีการนี้จำเป็นที่จะต้องมีคลังข้อมูลทางภาษา (Corpus) เป็นจำนวนมากและจะต้องมีการเรียนรู้การสร้างคำในบริบทหรือการเกิดร่วมกันของคำแต่ละคำเพื่อให้มีข้อมูลที่จะนำมาใช้ในการตัดคำ แต่เป็นวิธีที่สามารถแก้ไขปัญหาความกำกวมในการตัดคำ

ตารางที่ 2.5 เปรียบเทียบหลักการตัดคำ

หลักการตัดคำ	ข้อดี	ข้อเสีย
การใช้กฎ	- มีความรวดเร็วในการตัดคำ - ไม่ต้องการข้อมูลใน หน่วยความจำ ของคอมพิวเตอร์	- ยากที่จะสร้างกฎไวยากรณ์ได้ สมบูรณ์ - ทำได้ในระดับพยางค์ - ไม่สามารถจัดการคำที่คอมพิวเตอร์ ไม่รู้จัก (Unknown Words)
การใช้ พจนานุกรม	- มีความแม่นยำสูงในการตัดคำ - สามารถแก้ไขปัญหาคำที่ คอมพิวเตอร์ไม่รู้จัก (Unknown Words) โดยการเพิ่มเข้าไปใน พจนานุกรม	- ต้องการหน่วยความจำมากเพื่อจัดการ กับการเก็บพจนานุกรม - บริหารจัดการยาก ในการนำคำที่ คอมพิวเตอร์ไม่รู้จัก (Unknown Words) เข้าไปในพจนานุกรม
การใช้ คลังข้อมูล	- ไม่จำเป็นต้องเก็บพจนานุกรม ในหน่วยความจำ - แก้ไขปัญหาคำที่คอมพิวเตอร์ ไม่รู้จัก (Unknown Words) ได้ บางกรณี โดยการเรียนรู้จาก คลังข้อมูล - สร้างกฎไวยากรณ์ได้จากการ เรียนรู้ จากคลังข้อมูล	- ต้องการคลังข้อมูลที่มีขนาดใหญ่ เวลาในการเรียนรู้ และต้องการ วิธีการ (Algorithm) ที่ดีในการ เรียนรู้

ตารางที่ 2.6 สรุปคุณลักษณะข้อเด่น ข้อด้อย และตัวอย่างการนำไปใช้งานที่เหมาะสม

โปรแกรมตัดคำ (เทคนิคตัดคำ)	ข้อเด่น	ข้อด้อย	นำไปใช้
แบบย้อนรอยกลับ (Back Tracking)	ทำให้ได้จำนวนคำผิด น้อยที่สุด มีวิธีการกู้คืน จากความผิดพลาด	ทำให้ได้จำนวนคำ ทั้งหมดน้อยที่สุด	ตรวจสอบ คำสะกด
แบบเทียบคำยาว ที่สุด (Longest Matching)	ทำให้ได้จำนวนคำที่ ถูกต้องมาก ที่สุด	ทำให้เสียเวลาหากคำที่ ย้อนกลับไปแล้วยังไม่ พบตามพจนานุกรม	จัดรูปแบบ เอกสาร

ตารางที่ 2.6 (ต่อ)

โปรแกรมตัดคำ (เทคนิคตัดคำ)	ข้อเด่น	ข้อด้อย	นำไปใช้
แบบเทียบคำยาวที่สุด (Longest Matching)	ทำให้ได้จำนวนคำที่ ถูกต้องมากที่สุด	ทำให้เสียเวลาหากคำที่ย้อนกลับไปแล้วยังไม่ พบตามพจนานุกรม	วิเคราะห์ข้อความภาษาไทยในประโยค
แบบเทียบคำสั้นที่สุด (Shortest Matching)	ทำให้ได้จำนวนคำทั้งหมดมากที่สุด	ทำให้ได้จำนวนคำผิด มากที่สุด ได้จำนวนคำ ที่ถูกต้องน้อยที่สุด ได้สัดส่วนความถูกต้อง ต่อจำนวนคำในพจนานุกรมต่ำที่สุด	ไม่เหมาะกับการนำไปประยุกต์ใช้
แบบอาศัยความถี่คำ (Word Frequency)	ทำให้ได้สัดส่วนความถูกต้องต่อจำนวนคำในพจนานุกรมสูงที่สุด		ต้องไปพัฒนาเพิ่มเติม จึงจะนำไปใช้งานได้
การตัดคำให้ได้จำนวนคำและคำที่ไม่มีในพจนานุกรมน้อยที่สุด (Maximal - Matching)	-	-	วิเคราะห์ข้อความภาษาไทยในประโยค
แบบใช้พจนานุกรมคำกำกวม (Ambiguity Dictionary)	ทำให้ได้ความผิดพลาด ที่เกิดจากคำกำกวม น้อยที่สุด		สังเคราะห์เสียงพูดจากประโยคการแปลภาษา ด้วยเครื่อง

4. การวิเคราะห์ความรู้สึก (Sentiment Analysis)

การวิเคราะห์ความรู้สึก คือ การแยกแยะความรู้สึกจากคำที่อยู่ในประโยค ที่สามารถบอกได้ว่าความรู้สึกที่มีต่อสิ่งที่อยู่ในประโยคว่าเป็นความรู้สึกดี (ชอบ) หรือ ไม่ดี (ไม่ชอบ) เพื่อบ่งบอกความรู้สึกของผู้คนที่มีต่อบางสิ่งบางอย่าง ซึ่งการวิเคราะห์ความรู้สึกจะเกี่ยวข้องกับข้อความหรือประโยคที่แสดงถึงความรู้สึก ลักษณะของความรู้สึกว่า “ดี” หรือ “ไม่ดี” และความเกี่ยวข้องกันระหว่างความรู้สึกกับส่วนที่เป็นประธานของประโยค ซึ่งทั้งหมดจะมีความสัมพันธ์เกี่ยวข้องกัน ตัวอย่างของประโยคเช่น “สินค้า ก ดีกว่า สินค้า ข” จะเห็นได้ว่าคำว่า ดีกว่า เป็นตัวบ่งบอกถึงความรู้สึกเชิงบวก (ดี) ที่มีต่อสินค้า ก และยังส่งผลบ่งบอกถึงความรู้สึกเชิงลบ (ไม่ดี) ไปยังสินค้า ข ด้วย (นริศร์, 2550) (Tetsuya and Jeonghee, 2003) โดยการวิเคราะห์ความรู้สึกในประโยคจะใช้หลักการของการประมวลผลภาษาธรรมชาติ (NLP) ด้วยการทำให้เป็นรูปประโยคอย่างง่ายและวิเคราะห์จากชนิดของคำ (Part of Speech) โดยวิเคราะห์ในส่วนของคำที่อยู่ด้วยกันเป็นรูปแบบของประโยค โดยใช้กระบวนการแยกแยะข้อมูล (Information Extraction) เพื่อวิเคราะห์ส่วนที่แสดงถึงความสัมพันธ์ที่เกี่ยวข้องกัน การวิเคราะห์หาความสัมพันธ์กันขึ้นอยู่กับขอบเขตที่ระบุไว้ด้วย ตัวอย่างเช่น ประโยคที่มีประธานและส่วนที่แสดงความรู้สึกนั้นอยู่ในประโยคเดียวกันและยังอยู่ใกล้กัน แต่มีความรู้สึกที่แสดงออกมาทั้งสองส่วนกลับขัดแย้งกัน ตัวอย่างเช่น “ฉันชอบรองเท้าคู่นี้ใส่สบาย แต่สีไม่สวย” จะเห็นได้ว่า ภายในประโยคมีสองส่วนที่ขัดแย้งกัน คือ “ชอบรองเท้าเพราะใส่สบาย” แต่กลับไม่ชอบ“สีของรองเท้า” ซึ่งทำให้เป็นต้นเหตุสำคัญที่ทำให้การวิเคราะห์ความสัมพันธ์กันระหว่างประธานและส่วนที่บอกถึงความรู้สึกได้ผลลัพธ์ออกมาไม่ดีเท่าที่ควร ซึ่งอาจเกิดจากโปรแกรมคอมพิวเตอร์ที่ใช้ในการประมวลผลนั่นเอง โดยโปรแกรมส่วนใหญ่จะเน้นไปที่การพยายามวิเคราะห์ค่าของความพึงพอใจที่เป็นเชิงบวกหรือเป็นเชิงลบที่มีต่อสิ่งนั้นออกมาทั้งหมดในคราวเดียว ไม่ว่าจะสามารถวิเคราะห์ได้อย่างชัดเจนหรือไม่ชัดเจนก็ตาม ตัวอย่างเช่น การวิเคราะห์การรีวิวร้านอาหาร ว่าร้านนี้ดีหรือไม่ดี คือ อาจจะให้ค่าความรู้สึกที่เป็นลบ สำหรับพนักงานร้าน เช่น “พนักงานร้านนี้บริการได้แย่มาก” แต่ว่าในส่วนของอาหารทั้งหมดกลับให้ค่าเป็นบวก เช่น “อาหารร้านนี้ถือว่าอร่อยมาก” เป็นต้น ทำให้เกิดความสบสนในการตัดสินใจได้ในทางตรงกันข้าม ถ้าใช้การวิเคราะห์โดยใช้ความสัมพันธ์ระหว่างประธานและส่วนที่บอกถึง

ความรู้สึก สามารถวิเคราะห์ให้ละเอียดลงไปกว่านั้นได้ว่า อะไรที่ดีและอะไรที่ไม่ดีการกำหนดรูปแบบที่แสดงถึงความพึงพอใจนอกจากคำที่ขยายคำนามแล้วคำอื่น ๆ ก็สามารถนำมาใช้เพื่อบอกถึงความรู้สึกชอบหรือไม่ชอบได้ เช่น คำนาม คำกริยา และคำขยายที่บอกถึงความรู้สึกเช่นคำว่า “ดี” เป็นคำที่บอกความรู้สึกที่มีต่อคำนาม เช่น “สินค้าดี” และในประโยคของคำนามทั้งหมดจะเป็นการบอกถึงชนิดของความรู้สึกที่มีคำเดียวกับคำขยาย (“ดี” ในกรณีนี้มีค่าเป็นบวก) ในทำนองเดียวกันการบอกความพึงพอใจโดยใช้คำขยายกริยา เช่น คำว่า “มาก” เป็นตัวที่บอกถึงความพึงพอใจของคำกริยา เช่น “ทำงานเร็วมาก” โดยชนิดของความพึงพอใจจะได้มาจากคำกริยา ดังนั้นการบ่งบอกถึงความพึงพอใจ โดยใช้คำขยายคำนาม คำขยายกริยา และคำนาม สามารถบอกได้ว่าความพึงพอใจเป็นเชิงบวกหรือเป็นเชิงลบในทางกลับกัน ดังตัวอย่างที่ผ่านมา เช่น “สินค้า ก ดีกว่า สินค้า ข” ชนิดของความพึงพอใจที่ได้ออกมาจากคำกริยาจะต้องขึ้นอยู่กับความสัมพันธ์กันด้วย ในกรณีนี้ค่าความพึงพอใจที่เป็นเชิงบวกจะชี้ไปที่ประธาน (สินค้า ก) และค่าความพึงพอใจที่เป็นเชิงลบจะชี้ไปที่กรรม (สินค้า ข) ยิ่งไปกว่านั้น คำกริยาบางคำไม่สามารถที่จะบอกความพึงพอใจได้ด้วยตัวเอง แต่เป็นเพียงตัวที่ส่งผ่านค่าความพึงพอใจไปยังส่วนอื่นที่เกี่ยวข้อง เช่น คำว่า “มี” จะเป็นตัวส่งผ่านความพึงพอใจไปยังประธาน ตัวอย่างเช่น “สินค้ามีมาก” ซึ่งคำว่า “มาก” มีค่าความพึงพอใจที่เป็นเชิงบวก ซึ่งจะส่งผลไปถึงประธานทำให้รูปประโยคมีค่าความพึงพอใจเป็นเชิงบวกไปด้วย (นริศร์, 2550)

5. ตัวจำแนกข้อมูลนาอีฟ เบย์ (Naive Bayes Classifier)

การจำแนกข้อมูลเป็นกระบวนการสร้างตัวจัดการข้อมูล เพื่อแสดงให้เห็นความแตกต่างระหว่างกลุ่มของข้อมูล และเพื่อพยากรณ์ว่าข้อมูลนี้ควรจัดอยู่ในกลุ่มใด ซึ่งโมเดลที่ใช้จำแนกข้อมูลออกเป็นกลุ่มตามที่ได้กำหนดไว้จะขึ้นอยู่กับการวิเคราะห์ข้อมูลเรียนรู้ (Training Data) (บุญเสริม, 2002) อัลกอริทึมที่นิยมใช้ในการจำแนกข้อมูล คือ ตัวจำแนกข้อมูลนาอีฟ เบย์ (Naive Bayes Classifier)

ตัวจำแนกข้อมูลนาอีฟ เบย์ (Naive Bayes Classifier) เป็นการจำแนกข้อมูลที่ใช้หลักบนความน่าจะเป็นของข้อมูลบนพื้นฐานของทฤษฎีเบย์ (Bayes Theorem) และสมมติฐานที่กำหนดในการเกิดของเหตุการณ์ต่าง ๆ ที่ใช้ในการจำแนกนั้นเป็นอิสระต่อกัน ตัวจำแนกข้อมูลนาอีฟ เบย์

(Naive Bayes Classifier) เป็นตัวจำแนกข้อมูลที่มีประสิทธิภาพจึงมีการนำไปประยุกต์ใช้งานในด้านการจำแนกประเภทข้อความ (Text Classification) โดยมีสมการดังนี้

$$P(B|A) = \frac{P(A \cap B)}{P(B)} \quad (2\text{-}1)$$

โดยที่ P(B|A) คือ ค่าความน่าจะเป็นที่เกิดเหตุการณ์ B ขึ้นก่อนและจะมีเหตุการณ์ A ตามมา
P(A∩B) คือ ค่าความน่าจะเป็นที่เหตุการณ์ A และเหตุการณ์ B เกิดขึ้นร่วมกัน
P(B) คือ ค่าความน่าจะเป็นที่เหตุการณ์ B เกิดขึ้น

ในลักษณะเดียวกันสามารถเขียน P(B|A) หรือความน่าจะเป็นที่เหตุการณ์ A เกิดขึ้นก่อนและเหตุการณ์ B เกิดขึ้นตามมาทีหลังได้เป็น

$$P(A|B) = \frac{P(A \cap B)}{P(A)} \quad (2\text{-}2)$$

โดยที่ P(A|B) คือ ค่าความน่าจะเป็นที่เกิดเหตุการณ์ B ขึ้นก่อนและจะมีเหตุการณ์ A ตามมา
P(A∩B) คือ ค่าความน่าจะเป็นที่เหตุการณ์ A และเหตุการณ์ B เกิดขึ้นร่วมกัน
P(A) คือ ค่าความน่าจะเป็นที่เหตุการณ์ A เกิดขึ้น

จากทั้ง 2 แบบจะเห็นว่ามีค่า P(A∩B) ที่เหมือนกันอยู่ดังนั้นสามารถเขียนสมการของ P(A∩B)ได้เป็นดังนี้

$$P(A|B) = \frac{P(A|B) x P(B)}{P(A)} \quad (2\text{-}3)$$

โดยที่ P(A|B) คือ ค่าความน่าจะเป็นที่เกิดเหตุการณ์ B ขึ้นก่อนและจะมีเหตุการณ์ A ตามมา
P(B) คือ ค่าความน่าจะเป็นที่เหตุการณ์ B เกิดขึ้น
P(A) คือ ค่าความน่าจะเป็นที่เหตุการณ์ A เกิดขึ้น

งานวิจัยชิ้นนี้ได้นำสมการ Bayes Theorem หรือทฤษฎีของเบย์มาใช้งาน และเพื่อให้สอดคล้องและเข้าใจได้ง่ายขึ้น จึงเปลี่ยนสัญลักษณ์ A และ B ให้เป็น A และ C โดยที่ A คือ แอตทริบิวต์ (Attribute) และ C คือ ค่าคลาส (Class) ดังสมการด้านล่าง

$$\overbrace{P(C|A)}^{Posterior\ probability} = \frac{\overbrace{P(A|C)}^{Likelihood}\ x\ \overbrace{P(C)}^{Prior\ Probability}}{P(A)} \quad (2\text{-}4)$$

โดยที่ P(C|A) คือ ค่าความน่าจะเป็นที่ข้อมูลที่มีแอตทริบิวต์เป็น A จะมีคลาส C

P(A|C) คือ ค่าความน่าจะเป็นที่เซตข้อมูลทดสอบ (Training Data) ที่มีคลาส C และมีแอตทริบิวต์ A โดยที่ A = $a_1 \cap a_2 \ldots \cap a_m$ โดยที่ m คือจำนวนแอตทริบิวต์ในเซตข้อมูลทดสอบ (Training Data)

P(C) คือ ค่าความน่าจะเป็นของคลาส C

แต่การที่แอตทริบิวต์ A = $a_1 \cap a_2 \ldots \cap a_m$ ที่เกิดขึ้นในเซตข้อมูลทดสอบ (Training Data) อาจจะมีจำนวนน้อยมากหรือไม่มีรูปแบบของแอตทริบิวต์ (Attribute) แบบนี้เกิดขึ้นเลย ดังนั้นจึงจำเป็นต้องใช้หลักการ โดยให้แต่ละแอตทริบิวต์ (Attribute) เป็นอิสระต่อกัน ทำให้สามารถเปลี่ยนเป็นสมการ P(A|C) ได้ดังนี้

$$P(A|C) = P(a_1|C)\ x\ P(a_2|C)\ x \ldots P(a_m|C) \quad (2\text{-}5)$$

6. การแจ้งเตือนผ่านทาง Line ด้วย Line Messaging API

LINE Application นั้นเป็นโปรแกรมที่ถูกสร้างขึ้นในช่วงกลางปี 2010 โดยการร่วมมือของบริษัท Naver Japan Corporation และบริษัท Livedoor โดยมี NHN Japan เป็นผู้พัฒนาฟีเจอร์ต่าง ๆ ของไลน์ และในส่วนของการตลาดด้านธุรกิจนั้นยกให้บริษัทแม่ที่เกาหลี NHN Corporation จัดการ หลังจากที่เปิดตัวได้เพียงไม่นาน ก็ได้รับการตอบรับถึงหลายสิบล้านยูสเซอร์ในญี่ปุ่น ประเด็นแรกที่ใช้ในการสร้างโปรแกรมแชท LINE ขึ้นมาก็มีสาเหตุมาจากเหตุการณ์แผ่นดินไหวที่ภูมิภาค Tohoku เมื่อต้นปี 2011 ซึ่งในตอนนั้นระบบการติดต่อทางโทรศัพท์ล่ม ทำให้ NHN Japan ตัดสินใจออกแบบ App ที่สามารถใช้ได้ทั้งบนมือถือ บนแท็บเล็ต และคอมพิวเตอร์พีซี ซึ่งจะทำงานบนเครือข่ายข้อมูลที่สามารถแชทตอบโต้ได้รวดเร็วและต่อเนื่อง ด้วยความที่ไลน์มีคุณสมบัติของโปรแกรมแชทครบถ้วน ตั้งแต่ แชท ส่งไฟล์รูป ไฟล์วิดีโอ ไฟล์เสียง ระบบการค้นหา

เพื่อนด้วย QR Code หรือจะเล่นเกมสำหรับคลายเครียด ยังมีอีกสิ่งหนึ่งที่ถือได้ว่าเป็นจุดเด่นของแอปพลิเคชันนี้ก็คือ “Sticker”

แอปพลิเคชันไลน์ที่ผสมผสานบริการ Messaging และ Voice Over IP นำมาผนวกเข้าด้วยกัน จึงทำให้เกิดเป็นแอปพลิเคชันที่สามารถแชท สร้างกลุ่ม ส่งข้อความ โพสต์รูปต่าง ๆ หรือจะโทรคุยกันแบบเสียงก็ได้ โดยข้อมูลทั้งหมดไม่ต้องเสียเงิน สามารถใช้งานร่วมกันระหว่าง iOS และ Android รวมทั้งระบบปฏิบัติการอื่น ๆ ได้อีกด้วย การทำงานของ Line นั้น มีลักษณะคล้าย ๆ กับ WhatsApp ที่ต้องใช้เบอร์โทรศัพท์เพื่อยืนยันการใช้งาน แต่ Line ได้เพิ่มลูกเล่นอื่น ๆ เข้ามา ทำให้ Line มีจุดเด่นที่เหนือกว่า WhatsApp Line เป็นอีกหนึ่งตัวเลือกสำหรับแอปพลิเคชันสำหรับแชท ส่งข้อความ, เสียงและรูปภาพ ด้วยลูกเล่นของแอปพลิเคชันที่มีเสน่ห์การใช้งานที่ทำให้ผู้ใช้ชื่นชอบแถมทำงานได้อย่างดีเยี่ยม ยิ่งถ้าหากอยู่ในพื้นที่ที่อินเทอร์เน็ตครอบคลุม สามารถใช้ Line คุยกันทดแทนโทรศัพท์ได้เลย อีกทั้งยังเป็นแอปพลิเคชัน ที่แจกฟรี สามารถใช้งานได้บน iOS, Android และระบบปฏิบัติการอื่น ๆ

Line Messaging API คือ API (Application Programing Interface) ที่ Line พัฒนาขึ้นมาเพื่อให้นักพัฒนาระบบสามารถส่งข้อความจากระบบของตนเองไปยังไลน์ของผู้ใช้งานที่ได้ติดตามหรือเป็นเพื่อนอยู่ ซึ่งนักพัฒนานิยมนำมาใช้งานในการส่งข้อความแจ้งเตือน และพัฒนาเป็น Line Bot นักพัฒนาสามารถสมัครใช้งานได้ฟรีแบบมีเงื่อนไขรองรับการทำงานที่หลากหลายภาษาครอบคลุมทุกระบบปฏิบัติการ

การทำงานของ Messaging API (LINE API) Messaging API ทำการเชื่อมต่อระหว่าง User ผ่านทาง LINE Official Account หรือ LINE@ Account ซึ่งด้วย Messaging API นี้นักพัฒนาจะสามารถรับเพื่อน (Accept Friend) รวมถึงส่งข้อความหา User คนอื่น ๆ ที่เพิ่มบัญชีเป็นเพื่อน โดยผ่านหน้า LINE@ Manager ที่ตั้งไว้ หรือ ส่งออกจากจาก Server ก็ได้ในรูปแบบโต้ตอบ (Interactive)

การใช้งาน Messaging API ทำให้คุณสามารถส่งข้อมูลระหว่าง Server ไปยัง User LINE ผ่านทาง LINE Platform ซึ่งการ Request ที่ใช้ส่งข้อมูลต้องอยู่ในรูป JSON Format โดยตัว Server จะต้องเชื่อมต่อกับ LINE Platform และเมื่อ มี User เพิ่ม Account LINE เป็นเพื่อน หรือ ส่งข้อความมาหา ทาง LINE Platform จะทำการส่ง Request มายัง Server ที่ลงทะเบียนผูกไว้กับ LINE Account นั้นทันที วิธีนี้เรียกว่า Webhook ซึ่งจะทำให้ผู้ใช้งานรู้สึกเหมือนกับว่าได้โต้ตอบกับคน

การรับข้อมูลจาก LINE Platform ข้อมูลจะถูกส่งมาที่ URL ของ Server เมื่อไรก็ตามที่มี User ทำการติดต่อกับ Account ซึ่งมีด้วยกัน 2 แบบ คือ

1) User ส่งข้อความ (Message)

2) User มี Action เช่น Add LINE ID ขอเป็นเพื่อน (Operation)

รูปแบบข้อมูลที่ส่งมาจะผ่านทาง HTTP ที่ตั้ง URL ไว้กับ LINE โดยข้อมูลจะอยู่ในรูป JSON Format ซึ่งมีเนื้อข้อมูลตาม Operation ที่ User ติดต่อไป และทุก ๆ Request จะมี Signature ใส่ไปในส่วนของ Header ด้วย ซึ่ง Server จะต้องทำการตรวจสอบ Signature นี้ว่าถูกส่งมาจาก LINE Platform จริง ๆ ไม่ใช่มีผู้ปลอมแปลงส่งมา

การส่งข้อมูลไปหา LINE Platform สามารถใช้ API ที่ทาง LINE Platform ให้ไว้ใช้สำหรับการส่งข้อมูลจาก Server ส่วนตัวไปหา User ซึ่งความสามารถที่ทำได้คือ

1) ส่งข้อความ ไปยัง User หรือ ใครก็ตามที่เป็นเพื่อน

2) ดึงข้อมูลชื่อ Display ของ User

ซึ่ง Server สามารถเรียก API เพื่อส่งข้อ Message ไปหา User ได้ตลอดเวลา โดยจะต้องทำการตั้งค่า Channel Access Token (Channel ID), Channel Secret และ Channel MID ซึ่ง API ทั้งหมดนี้ต้องใช้ผ่าน HTTPS

Messaging API Specifications วิธีใช้ Line API ทั้งหมด จะใช้ผ่าน Http/Https Protocol โดยใช้ Post Method และ ใช้ข้อมูล Body Message เป็น Json Format เป็นตัวส่ง Request ไปหา Line Server ซึ่งอาจจะทำเป็น Restful Webservice

Rate Limitsในการใช้งาน LINE API ทุกชนิดจะมีข้อจำกัด หรือ Limit ที่ทาง LINE Set ไว้ตามระดับของ Id ที่สมัครไว้ ซึ่งโดยปกติแล้วของนักพัฒนาจะเป็น Developer Trial โดยจะสามารถส่ง API ไปยัง LINE Server ได้สูงสุดไม่เกิน 1000 Request/Min ตามตารางที่ 2.7

ตารางที่ 2.7 อัตราการใช้บริการส่งข้อความด้วย LINE API

Plan	**Limit**
Developer Trial	1,000/Min
Other Plans	10,000/Min

Status Codes คือผลลัพธ์การร้องข้อข้อมูล (Response) จากฝั่ง LINE ที่ตอบกลับมาหา เพื่อแจ้งผลลัพธ์ของการร้องขอข้อมูล (Request) ที่ส่งข้อมูลประเภท Http POST ไป ดังตารางที่ 2.8

ตารางที่ 2.8 ข้อมูลการตอบกลับจากเซิฟเวอร์ไลน์

Error Code	**Description**
200 OK	ถูกต้องตามกฎ
400 Bad Request	ข้อมูลที่ส่งไปผิด Format Schema
401 Unauthorized	Channel Access Token ไม่ถูกต้อง
403 Forbidden	ฝั่ง Server ปฏิเสธ Request เนื่องจาก Account Id ไม่ได้รับอนุญาตในการใช้งานคำสั่งนั้น
429 Too Many Requests	Request มากเกิน Limit ของ Account Plan
500 Internal Server Error	Server LINE มีปัญหา

Response Headers กรณีที่ได้รับ Response Status Code 200 Ok ทาง Line Server แปะ Request-Id กลับมาให้ ซึ่งจะไม่ซ้ำกันดังตารางที่ 2.9

ตารางที่ 2.9 ข้อมูล Header การตอบกลับจากเซิฟเวอร์ไลน์

Response Header	**Description**
X-Line-Request-Id	Generate จากฝั่ง LINE Server

Error Response กรณีที่ได้รับ Response Status Code 200 OK แต่ทาง LINE Server ไม่สามารถ Process คำสั่งนั้นได้ อาจจะเพราะข้อมูลไม่ครบถ้วน หรือ คำสั่งติดเงื่อนไขบางอย่าง ทาง LINE Server จะตอบ Error กลับมาพร้อมรายละเอียดดังตารางที่ 2.10

ตารางที่ 2.10 ข้อมูลสถานการณ์ตอบกลับในกรณีที่เกิดข้อผิดพลาดจากเซิฟเวอร์ไลน์

Field	Type	Description
Message	String	บอกจำนวนที่ Error
Details[].Message	String	รายละเอียดของ Error

Messaging API สำหรับคำสั่ง (Command) ที่ทาง LINE ออกแบบมาให้ใช้งาน มีส่วนประกอบดังนี้

1) **Webhooks** ใช้รับการแจ้งเตือน (Notification) ที่เกิดขึ้นกับ Line Account แบบ Real-Time สามารถเอามา ทำ Line Bot รับข้อความ (Message)ได้
2) **Reply Message** ตอบข้อความกลับไปหา User ที่ส่งข้อความมา
3) **Push Message** ส่งข้อความหา User ได้ตลอดเวลา
4) **Multicast** ส่งข้อความหาหลาย ๆ User พร้อมกัน (Broadcast)
5) **Content** ดาวน์โหลด รูป Video และ ข้อความเสียงที่ User ส่งมา
6) **Profile** ดึงข้อมูล User Profile
7) **Leave** ออกจากกลุ่มไลน์หรือห้องสนทนา

7. ระบบเรียลไทม์ (Real-time System)

ระบบเรียลไทม์ (Real-time) คือระบบที่สามารถให้การตอบสนองจากระบบอย่างทันทีทันใดเมื่อได้รับอินพุตเข้าไป ในทางอุดมคติระบบเรียลไทม์นี้จะเป็นระบบที่ไม่เสียเวลาในการประมวลผลหรืออาจจะกล่าวได้ว่าเวลาในการประมวลเป็นศูนย์ แต่ในทางปฏิบัติเครื่องคอมพิวเตอร์ที่ทำงานแบบเรียลไทม์นี้ไม่สามารถผลิตขึ้นมาได้ ทำได้เพียงการลดเวลาการประมวลผลให้น้อยที่สุด จนไม่สามารถเห็นความแตกต่างของช่วงเวลาที่ป้อนอินพุตเข้าไปและได้รับเอาต์พุตออกมา เวลาของความแตกต่างนี้เรียกว่า “เวลาตอบสนอง” (Response Time) ซึ่งผู้ใช้งานทั่วไปต้องการเวลาตอบสนองให้น้อยที่สุดเพื่อประสิทธิภาพของระบบ ระบบเรียลไทม์นิยมนำไปใช้ในการควบคุมกระบวนการทางอุตสาหกรรม ซึ่งปัจจุบันสามารถควบคุมด้วยระบบคอมพิวเตอร์ที่ทันสมัยให้ค่าเวลาตอบสนองที่ยอมรับได้ สำหรับในวงการคอมพิวเตอร์ระบบเรียลไทม์เข้าใกล้อุดมคติมากขึ้น

เนื่องจากความเร็วในการประมวลผลของซีพียูที่เมื่อสั่งให้โปรแกรมทำงานแล้วต้องคำนวณผลลัพธ์ออกมาทันทีไม่ใช่รออีก หลายนาที Real-time Programming คือ Program ที่ตอบสนองแสดงเวลาสถานะของเครื่องจริง ระบบเรียลไทม์ (Real-time system) ระบบปฏิบัติการ คือ ระบบเวลาจริง (Real-time system) หมายถึงการตอบสนองทันที เช่นระบบ Sensor ที่ส่งข้อมูลให้คอมพิวเตอร์ เครื่องมือทดลองทางวิทยาศาสตร์ ระบบภาพทางการแพทย์ ระบบควบคุมในโรงงานอุตสาหกรรม ระบบหัวฉีดในรถยนต์ ระบบแขนกล และเครื่องใช้ในครัวเรือนทั้งหมดซึ่งแม้แต่หน้าไทม์ไลน์ของ Windows 8 ก็เป็นระบบ Real time

Web Socket คือการสื่อสารระหว่าง Web Server กับ Client แบบ เรียลไทม์ (Real Time) Web Socket สำหรับนักพัฒนาที่เคยเขียนโปรแกรมแบบ Client-Server นั้น จะคุ้นเคยกับคำว่า Server Socket ซึ่งเป็นการเปิด Port ที่ Server เป็น TCP ซึ่ง Client ก็ติดต่อโดยอ้างถึง IP ของ Server และ Port และทำการเชื่อมการติดต่อเพิ่มเริ่มส่งข้อมูล Web Socket ซึ่งเป็นเทคโนโลยีเพื่อใช้การติดต่อสื่อสารระหว่าง Web server กับ Client แบบ Real Time Client ที่เป็น Html5 + Javascript สามารถรับข้อมูลทางฝั่ง Server มาแสดงผลได้ ผ่าน Protocol TCP/IP โดยไม่ต้อง Refresh หน้า เช่นเดียวกับ Ajax แต่ Websocket นั้นมีข้อดีกว่า Ajax คือ Websocket ไม่ต้องส่ง Request ใหม่ เพื่อส่งคำร้องขอไปยัง Server และ รอรับ Respond จากทางฝั่ง Server ทำให้ Websocket ประหยัดทั้งเวลา และปริมาณข้อมูลที่ส่งไปมาระหว่าง Web Server กับ Client โดยการทำงานของ Websocket มันจะรอรับ Message ทางฝั่ง Web Server อยู่ตลอดเวลา หลังจากที่ได้ส่งคำสั่ง ติดต่อ Websocket Server ไปแล้วในครั้งแรก และมันจะรอรับ Message จนกระทั้งจะสั่ง Close หรือหยุดการติดต่อ ตัวอย่างโค้ด Websocket ดังภาพที่ 2.3

```
var socket = new WebSocket(ws://websockets.org:8787/echo);
socket.onopen = function (evt) {
   console.log("Socket opened");
};
socket.onclose = function (evt) {
   console.log("Socket closed");
};
socket.onmessage = function (evt) {
   console.log(evt.data);
};
socket.onerror = function (evt) {
   console.log("Error: " + evt.data);
};
socket.send("Hello World!");
```

ภาพที่ 2.3 ตัวอย่างโค้ด Websocket

จากภาพที่ 2.3 เป็นการสร้าง Websocket เพื่อไว้สำหรับติดต่อกับ Echo Server และมีการกำหนด Callback Function เมื่อเกิดเหตุการณ์ต่างๆ เช่น Opened, Closed, Receive A Message, หรือแม้กระทั่งเมื่อเกิด Error ขึ้น จากนั้นระบบจะส่งค่า Hello World! ไปยัง Server และให้ Browser แสดงคำว่า Hello World จากการรับค่ากลับจาก Server ซึ่ง Server ของ Websocket หากลองค้นหาดูแล้ว จะพบว่ามีการ Implement เป็นหลายภาษาอย่างแพร่หลายแล้ว นอกจากนี้ยังสามารถทำงานร่วมกับ HTML อีกด้วย

8. การทดสอบประสิทธิภาพ (Performance Evaluation)

การประเมินผลและการทดสอบเพื่อประเมินประสิทธิภาพการจำแนกข้อมูลในงานวิจัยนี้ จะใช้การวัด ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าความถูกต้อง (Accuracy) และการวัดประสิทธิภาพโดยรวม (F-measure) เพื่อทดสอบความถูกต้องโดยความหมายของแต่ละค่าที่นำมาทดสอบประสิทธิภาพมีดังต่อไปนี้ ซึ่งสามารถคำนวณได้จากสมการที่ 2-6 ถึงสมการที่ 2-9

ค่าความแม่นยำ (Precision)

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{2-6}$$

โดยที่ Precision คือ ค่าความแม่นยำ

True Positive (TP) คือ ค่าผลการทำนายทายได้ถูกต้อง (Actual) ที่เป็น Positive

False Positive (FP) คือ ค่าผลการทำนายไม่ถูกต้อง ตามค่าที่คาดหวัง ที่เป็น Negative

กรณีที่ต้องการค่าความแม่นยำถูกต้อง 100% จำเป็นที่ต้องให้ค่า FP ที่ทายผิดแบบเสียหายเชิงคุณภาพให้มีค่าเท่ากับศูนย์ ในความเป็นจริงค่าความแม่นยำ (Precision) ต้องเป็นเท่าไร ในขณะที่บางงานที่ต้องการความถูกต้อง 100% จำเป็นต้องออกแบบในการสร้างโมเดลให้ดี จำเป็นต้องใช้เวลาในการพัฒนานาน และต้องออกแบบอัลกอริทึม (Algorithm) ให้มีความเหมาะสมกับเนื้องาน เช่นผลการทำนายที่เกี่ยวข้องกับชีวิต ในขณะที่บางงานอาจจะมีการกำหนดเปอร์เซ็นต์ข้อผิดพลาด (%Error) ที่สามารถยอมรับได้ แต่ในทางปฏิบัติอาจจะต้องเอาผลการทำงานมาตรวจสอบและใช้ค่าบ้างค่าในลักษณะ (Feature) หนึ่งมาเป็นตัวกลั่นกรอง สำหรับการเลือกรายการนั้นมาตรวจสอบซ้ำด้วยคนอีกครั้ง

ค่าเรียกคืน (Recall)

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{2-7}$$

โดยที่ Recall คือ ค่าเรียกคืน

True Positive (TP) คือ ค่าผลการทำนายทายได้ถูกต้อง (Actual) ที่เป็น Positive

False Negative (FN) คือ ค่าผลการทำนายไม่ถูกต้องตามที่คาดหวังที่เป็น Negative

กรณีที่ต้องการค่าเรียกคืนมีความถูกต้อง 100% จำเป็นต้องให้ค่า False Negative ที่ทายผิดแบบเสียหายเชิงปริมาณมีค่าเป็น ศูนย์ แต่ถ้าหากค่าเรียกคืน (Recall) มีค่าเท่ากับ 95% นั่นหมายความว่าจะมีรายการที่ทำนายผิด 5% ตัวอย่างเช่นความเป็นจริง (Actual) ผู้ป่วยไม่ได้เป็นมะเร็ง แต่ผลการทำนายออกมาว่าเป็นมะเร็งซึ่งไม่ถูกต้อง ในการแปลความหมายในลักษณะนี้จะเป็นการโยนผลดี เข้าไปกองรวมกับผลเสีย ซึ่งจากผลของการทำนาย ทำให้มีโอกาสเกิดความเสียหายได้สูงเชิงปริมาณ โดยความจริงคือผู้ป่วยไม่ได้เป็นมะเร็ง แต่ผลทำนายว่าเป็นมะเร็ง

ค่าความถูกต้อง (Accuracy)

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Data\ Set} \tag{2-8}$$

โดยที่ Accuracy คือ ค่าความถูกต้อง

True Positive (TP) คือ ค่าผลการทำนายทายได้ถูกต้อง (Actual) ที่เป็น Positive

True Negative (TN) คือ ค่าผลการทำนายทายได้ถูกต้องตามค่าที่คาดหวังที่เป็น Negative

Total Data Set คือ จำนวนข้อมูลทั้งหมดที่ใช้ในการทดสอบ

คือค่าที่บ่งบอกถึงความสามารถของเครื่องมือวัด (Instrument) ในการอ่านค่าหรือแสดงค่าที่วัดได้เข้าใกล้ค่าความเป็นจริง การหาอัตราส่วนของการทำนายผลที่ถูกต้องตามจำนวนข้อมูลทั้งหมด โดยการนำผลลัทธ์ค่าผลการทำนายได้ถูกต้อง TP (True Positive) รวมกับค่าผลการทำนายได้ถูกต้องตามค่าที่คาดหวังที่เป็น TN (True Negative) หารด้วยจำนวนข้อมูลทั้งหมดที่ใช้ในการทดสอบ

ค่าการวัดประสิทธิภาพโดยรวม (F-measure)

$$F\text{-}measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{2-9}$$

ค่าการวัดประสิทธิภาพโดยรวม (F-measure) หมายถึงค่าการวัดประสิทธิภาพโดยรวมของทั้งสองค่าระหว่างค่าความแม่นยำ และค่าเรียกคืนซึ่งนำค่าทั้งสองมาคำนวณร่วมกัน

9. งานวิจัยที่เกี่ยวข้อง

ปรีดาวรรณ เกษเมธีการุณ (2559)ได้ศึกษาการวิเคราะห์ความรู้สึกจากวีดีทัศน์บทวิจารณ์ภาษาไทยโดยใช้เทคนิคการทำเหมือง ข้อมูลโดยทำการเก็บรวบรวมข้อมูลความคิดเห็นในหัวข้อที่เกี่ยวกับผลิตภัณฑ์ความงามจาก Youtube นำมาสกัดเสียงและรู้จำเสียงอัตโนมัติการตัดคำ การเปรียบเทียบและวิเคราะห์ความรู้สึกการปรับปรุง คำคุณลักษณะการพัฒนาระบบการวิเคราะห์ความรู้สึกจากวีดีทัศน์บทวิจารณ์ภาษาไทย การประเมินประสิทธิภาพด้านความถูกต้อง การประเมินประสิทธิภาพด้านความพึงพอใจ การสร้างตัวแทนเอกสาร การสร้างแบบจำลองการวิเคราะห์ความรู้สึก และการเปรียบเทียบประสิทธิภาพของเทคนิคการเรียน ของเครื่อง (Machine Learning Technique) ผลการวิจัยพบว่าการปรับปรุงคำคุณลักษณะช่วยสนับสนุน ในการวิจัยชิ้นนี้ได้ใช้ การเชื่อมต่อกับโปรแกรม Thsplitlib ในการตัดคำ และการเชื่อมต่อฐานข้อมูล Mysql ซึ่งใช้เก็บคลังคำศัพท์คุณลักษณะ

ในงานวิจัยชิ้นได้เลือกใช้เทคนิคการตัดคำด้วย Thsplitlib ซึ่งงานวิจัยของปรีดาวรรณชี้ให้เห็นว่า Thsplilib เป็นเครื่องมือที่ใช้ในการตัดคำภาษาไทยที่มีประสิทธิภาพ มีความแม่นยำ และรวดเร็ว โดยถูกออกแบบมาเพื่อการตัดคำภาษาไทยโดยเฉพาะ ด้วยวิธีการเทียบคำกับฐานข้อมูลคลังคำศัพท์ (Corpus) โดยใช้การจับคู่ที่ยาวที่สุด (Longest Matching)

นริศร์ พรหมบุตร (2550) ได้ศึกษาการทำไมน์นิ่งความคิดเห็นในสินค้า โดยได้พัฒนาเป็นระบบในการวิเคราะห์ความคิดเห็นที่มีต่อสินค้า (Opinion Analysis) ประเภทโทรศัพท์มือถือของผู้ใช้งานจากเว็บไซต์
แล้วนำเสนอเทคนิคการวิเคราะห์ความรู้สึกด้วยกฎ (Rule-based Sentimental Analysis) สำหรับภาษาไทย โดยกฎที่สร้างขึ้นอาศัยลักษณะของสินค้า (Product's Features) ร่วมกับรูปแบบของประโยคภาษาไทย (Thai Sentence Pattern) ในงานวิจัยนี้ ผู้วิจัยได้ใช้วิธีการตัดคำแบบเปรียบเทียบคำกับพจนานุกรม และวิธีการตัดคำแบบการเลือกคำที่ยาวที่สุด (Longest Matching) แล้วทำการกำกับชนิดของคำในประโยค เพื่อค้นหาชนิดของคำ (Part Of Speech) แล้วทำการระบุคำที่เป็นคุณลักษณะของโทรศัพท์มือถือและคำที่บ่งบอกถึงความพึงพอใจที่จะถูกระบุเอาไว้ในฐานข้อมูล แล้วจะถูกนำไปตรวจสอบรูปแบบประโยคกับกฎที่กำหนดขึ้นโดยอิงตามหลักภาษาไทย เพื่อหาค่า

ความรู้สึกของผู้เขียนต่อลักษณะของสินค้านั้นว่าเป็นเชิงบวก เชิงลบ หรือเป็นกลาง และทำการประเมินโดยการเปรียบเทียบจากผลการวิเคราะห์ของคน พบว่าระบบมีประสิทธิภาพในการวิเคราะห์ด้านความแม่นยำ (Precision) เป็น 63% และค่าเรียกคืน (Recall) เป็น 74%

ในงานวิจัยชิ้นนี้ได้อ้างอิงทฤษฎีการตัดคำด้วยวิธีการตัดคำแบบการเลือกคำที่ยาวที่สุด (Longest Matching) มาใช้ โดยในงานวิจัยของ นริศร์ ชี้ให้เห็นว่าเทคนิคการตัดคำนี้เหมาะกับการตัดคำที่เป็นภาษาไทยและคำที่มาจากอินเทอร์เน็ตซึ่งมีความหลากหลาย

พนิดา ทรงรัมย์ (2559) ได้ศึกษาการจำแนกความคิดเห็นทางการเมืองบนเครือข่ายสังคมออนไลน์ โดยใช้วิธีการจำแนกแบบสัมพันธ์ ซึ่งเก็บรวบรวมข้อความความคิดเห็นบนเฟซบุ๊กเกี่ยวกับการดำเนินงานของ คสช ระหว่างวันที่ 22 พฤษภาคม 2557 ถึง วันที่ 17 กรกฎาคม 2557 โดยการใช้ Facebook Graph API ร่วมกับภาษา PHP เพื่อรวบรวมข้อความ โดยมีข้อมูลที่รวบรวมได้ทั้งหมด 467 ข้อความ นำข้อมูลไปจำแนก คือ ความคิดเห็นเชิงบวก และ ความคิดเห็นเชิงลบ โดยใช้วิธีการจำแนกแบบความสัมพันธ์ของข้อมูลที่เกิดร่วมกันบ่อยด้วยขั้นตอนวิธี CBA (Classification Based on Associations) และใช้ 10-fold Cross Validation ในการแบ่งข้อมูลเรียนรู้และข้อมูลทดสอบ ผลจากการจำแนกได้ความถูกต้อง 77.75%

ในงานวิจัยชิ้นนี้ได้ใช้เทคนิคการรวมรวมข้อมูลจากเฟซบุ๊กด้วย Facebook Grap API ซึ่งงานวิจัยของพนิดา ชี้ให้เห็นว่าเป็นกระบวนการเดียวที่จะสามารถสืบค้นข้อมูลจากเฟซบุ๊คได้ อีกทั้งงานวิจัยของพนิดาก็ได้พัฒนาขึ้นมาบนภาษา PHP ซึ่งมีความสอดคล้องกับงานวิจัยชิ้นนี้

Vishwasa Navada (2015) ได้ศึกษาการสกัดความรู้สึกแบบเรียลไทม์ และวิเคราะห์ข้อมูลสำหรับช่วยในการตัดสินใจ งานวิจัยชิ้นนี้ได้ศึกษาโดยการนำข้อความจากสื่อโซเชียลมีเดียคือ Twitter มาทำการวิเคราะห์ด้วยเทคนิคนาอีฟเบย์ แบบเรียลไทม์ โดยจำแนกผลการวิเคราะห์ออกเป็นสามส่วนคือ เป็นบวก (Positive) เป็นลบ (Negative) และเป็นกลาง (Neutral) ผู้วิจัยให้เหตุผลสำหรับการใช้เทคนิคนี้ว่า เป็นเทคนิคที่ใช้งานง่ายและให้ประสิทธิภาพที่ดี ผลลัพธ์ของการวิจัยคือ สร้างเครื่องมือช่วยในการตัดสินใจ ในการทำการตลาด เช่นการทำโฆษณา โดยแสดงข้อมูลในรูปแบบกราฟ

ในงานวิจัยชิ้นนี้ได้มีความสอดคล้องกับงานวิจัยของ Vishwasa ซึ่งเป็นการรวบรวมข้อมูลจากสื่อสังคมออนไลน์เหมือนกัน แต่ต่างกันที่แหล่งข้อมูล ซึ่งในงานวิจัยชิ้นนี้เลือกแหล่งข้อมูลที่เป็นเฟซบุ๊ก และกระบวนการวิเคราะห์ความรู้สึก Vishwasa ได้ใช้เทคนิคนาอีฟเบย์ (Naive Bayes Classification)

Muhammad Awis Jamaluddin Johari (2015) ได้ศึกษาการวิเคราะห์ความรู้สึกจากข้อความ โดยใช้เครื่องมือ Open Source ชื่อว่า Phpinsight ซึ่งเป็นเครื่องมือที่ใช้ในการวิเคราะห์ความรู้สึกแบบอัตโนมัติ ด้วยเทคนิคการจำแนกข้อมูลนาอีฟเบย์ Phpinsight สามารถวิเคราะห์ความรู้สึกได้ไว้ที่สุดเมื่อเปรียบเทียบกับ Textblob, Natural Language Tool Kit (NLTK), Alchemy และ Repustate งานวิจัยชิ้นนี้มุ่งเน้นการพัฒนาระบบวิเคราะห์ความรู้สึกแบบอัตโนมัติ เพื่อคำนวณหาค่าความถูกต้องที่ดีที่สุด โดยใช้ข้อมูลของบทความข่าวแบบเป็นทางการจำนวน 100 บทความ และแบ่งข้อมูลออกเป็น 10 หมวดหมู่ ผลการวิจัยพบว่า Phpinsight เป็นเครื่องมือที่ง่ายและมีประสิทธิภาพในการวิเคราะห์ข้อมูล

ในงานวิจัยชิ้นนี้ได้เลือกใช้เครื่องมือ Phpinsight โดยการนำมาประยุกต์ก่อนใช้งานคือ เพิ่มเติมในส่วนของการตัดคำภาษาไทยควบคู่ไปกับ การจัดการคำศัพท์ เดิมทีที่อยู่ในรูปแบบไฟล์ ให้อยู่ในรูปแบบฐานข้อมูล ในงานวิจัยของ Muhammad ชี้ให้เห็นว่า Phpinsight เป็นเครื่องมือที่พัฒนามาจากเทคนิการจำแนกข้อมูลนาอีฟเบย์ ซึ่งมีประสิทธิภาพสูงสุดเมื่อเทียบกับ Textblob, Natural Language Tool Kit (NLTK), Alchemy และ Repustate

Vivek Narayanan (2013) ได้ศึกษาการจำแนกความรู้สึกที่รวดเร็วและถูกต้องโดยใช้แบบเทคนิคการจำแนกข้อมูลนาอีฟเบย์ ขั้นสูง ในงานวิจัยชิ้นนี้ต้องการจะพัฒนาเพื่อให้หลักการวิเคราะห์ข้อมูลด้วยเทคนิคนาอีฟเบย์ นั้นมีค่าความถูกต้องมากที่สุดและเร็วที่สุด โดยการใช้สองหลักการในการวิเคราะห์คือ Word N-Gram ร่วมกับ Feature Selection เพื่อสร้างผลลัพธ์การคำนวณที่เร็วและถูกต้องมากที่สุด ผลการวิจัยได้ค่าความถูกต้องอยู่ที่ 88.80 %

ในงานวิจัยชิ้นนี้ได้ใช้เทคนิคการจำแนกข้อมูลนาอีฟเบย์ซึ่งผู้วิจัยได้เลือกใช้สมการในการวิเคราะห์ข้อมูลทั้งเอกสาร ซึ่ง Vivek ได้ชี้เห็นว่าเป็นเทคนิคที่เหมาะกับงานด้านการจำแนกข้อมูลทางเอกสาร (Text Classification) ด้วยเทคนิคนาอีฟเบย์

ชูชาติ หฤไชยะศักดิ์ (2561) ได้วิจัยและพัฒนาเฟรมเวิร์ค (Framework) การวิเคราะห์ความรู้สึกสำหรับโซเชียลมีเดีย (S-Sense) ซึ่งเป็นเครื่องมือสำหรับรวมรวม ติดตาม วิเคราะห์ ประมวลผลข้อความบนเว็บไซต์เครือข่ายเชิงสังคมและเว็บบอร์ด เช่น Facebook, Twitter, YouTube, Pantip.com เป็นต้น เนื่องจากข้อความส่วนใหญ่บนอินเทอร์เน็ตนิยมใช้ภาษาที่ไม่เป็นทางการหรือภาษาพูด รวมทั้งมีโครงสร้างประโยคที่ไม่ถูกต้องตามหลักไวยากรณ์ภาษาไทย จึงทำ

ให้ยากต่อการวิเคราะห์ ดังนั้นเทคโนโลยีของ S-Sense จึงถูกคิดค้นและพัฒนาขึ้นมาด้วยเทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing) และการทำเหมืองข้อความ (Text Mining) เพื่อให้สามารถวิเคราะห์ข้อความที่ใช้ภาษาพูดและไม่เป็นทางการได้อย่างมีประสิทธิภาพ ระบบ S-Sense สามารถนำมาช่วยให้ธุรกิจหรือองค์กรรับรู้กิจกรรมความเคลื่อนไหวต่างๆที่เกี่ยวข้องกับองค์กรของตนและตรวจสอบความพึงพอใจและติดตามทัศนคติของสาธารณะที่มีต่อยี่ห้อ ผลิตภัณฑ์ หรือบริการต่างๆของตนได้ เพื่อให้เข้าใจถึงความต้องการที่แท้จริงของลูกค้าได้ดียิ่งขึ้น ในงานวิจัยชิ้นนี้ได้ใช้เทคนิคการตัดคำแบบอิงพจนานุกรม (Dictionary Based) โดยใช้เทคนิคการแบ่งคำแบบ Longest Matching สำหรับการวิเคราะห์ความรู้สึกในงานวิจัยชิ้นนี้ได้ใช้เทคนิค Multinomial Naive Bayes Algorithm เหตุผลที่เลือกใช้เทคนิคเพราะว่า เป็นเทคนิคที่ต้องการจำนวนข้อมูลที่ใช้ในชุดทดสอบน้อย ผลลัพธ์สามารถวิเคราะห์และอธิบายได้อย่างง่าย

งานวิจัยชิ้นนี้ได้ใช้สองทฤษฎีหลักคือ การตัดคำภาษาไทยโดยใช้การจับคู่ที่ยาวที่สุด (Longest Matching) ซึ่งงานวิจัยของชูชาติและคณะได้ชี้ให้เห็นว่ามีประสิทธิภาพ รวดเร็วและถูกต้อง และการใช้การจำแนกข้อมูลด้วยเทคนิคนาอีฟเบย์ เป็นเทคนิคที่เหมาะกับข้อมูลที่มีปริมาณมาก มาจากแหล่งข้อมูลจากสังคมออนไลน์ (Social Media) ซึ่งการประมวลผลข้อมูลจำเป็นต้องมีทั้งความแม่นยำและความถูกต้อง และได้ถูกพิสูจน์แล้วโดยการนำไปใช้งานจริงและเป็นที่ยอมรับทั้งในองค์กรภาครัฐและเอกชน

จากการศึกษางานวิจัยที่เกี่ยวข้อง การวิเคราะห์ความรู้สึกบนสื่อสังคมออนไลน์ เทคนิคที่ใช้ในการจำแนกความรู้สึกที่ให้ประสิทธิที่ดีคือ เทคนิคการจำแนกข้อมูลด้วยนาอีฟเบย์ ซึ่งงานวิจัยของชูชาติและคณะเป็นตัวอย่างที่ดี เหมาะสำหรับการจำแนกข้อมูลประเภทข้อความ (Text Classification) ได้เป็นอย่างดี ในการเลือกเทคนิคมาใช้ในการจำแนกข้อมูลนั้นผู้วิจัยได้เลือกจากงานวิจัยผ่านการใช้งานจริงมาแล้วและเป็นที่ยอมรับ โดยคณะผู้วิจัยได้ระบุเอาไว้ว่าเทคนิคการจำแนกข้อมูลนาอีฟเบย์ นี้ต้องการจำนวนข้อมูลที่ใช้ในชุดทดสอบน้อย ผลลัพธ์สามารถวิเคราะห์และอธิบายได้อย่างง่าย อีกทั้งงานวิจัยของชูชาติและคณะยังเป็นที่ยอมรับยอมรับทั้งในองค์กรภาครัฐและเอกชน งานวิจัยชิ้นนี้จึงเลือกเทคนิคการจำแนกข้อมูลด้วยนาอีฟเบย์ มาใช้ในงานวิจัย อีกทั้งผู้วิจัยยังเลือกใช้เครื่องมือ Phpinsight ซึ่งเป็น Open Source ที่ผู้วิจัยนำมาพัฒนาต่อยอดบนภาษา PHP โดยออกแบบและพัฒนาให้สามารถเพิ่มคำศัพท์ใหม่ได้ตลอดเวลา ทั้งนี้ก็เพื่อเพิ่มประสิทธิภาพ

ในการทำงานที่ดีที่สุด นอกจากนี้ Phpinsight ยังเป็นเครื่องมือที่พัฒนาขึ้นมาโดยใช้เทคนิคการจำแนกข้อมูลนาอีฟเบย์ ที่มีประสิทธิภาพสูงสุดเมื่อเทียบกับ Textblob, Natural Language Tool Kit (NLTK), Alchemy และ Repustate จากงานวิจัยของ Muhammad

บทที่ 3
วิธีการดำเนินงานวิจัย

วิธีการดำเนินงานวิจัยในการพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ (Real-time) ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูล นาอีฟ เบย์ (Naive Bayes Classifier) สำหรับภาษาไทยสามารถแบ่งขั้นตอนการดำเนินงานออกเป็นดังนี้

1. การศึกษาและวิเคราะห์ปัญหา
2. การวิเคราะห์และออกแบบขั้นตอนการทำงานของระบบ
3. ข้อมูลและกลุ่มตัวอย่างที่ใช้ในงานวิจัย
4. การพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์
5. การทดสอบประสิทธิภาพของระบบ

1. การศึกษาและวิเคราะห์ปัญหา

จากการจัดอันดับมหาวิทยาลัย โดยสำนักงานคณะกรรมการอุดมศึกษา เมื่อปีพ.ศ. 2549 จึงทำให้การศึกษาระดับอุดมศึกษามีการแข่งขันกันมากขึ้น เพื่อเป็นการสร้างแรงจูงใจให้นักเรียนที่กำลังจะสำเร็จการศึกษาได้มีโอกาส และมีทางเลือกในการตัดสินใจเลือกสถานศึกษามากขึ้น ดังนั้น มหาวิทยาลัย จึงมีความจำเป็นอย่างยิ่งที่จะต้องมีกลยุทธ์ในการหาวิธีการที่จะให้โรงเรียนและนักเรียนกลุ่มเป้าหมายตัดสินใจเลือกเข้าศึกษาต่อระดับปริญญาตรีในมหาวิทยาลัย โดยการจัดกิจกรรมต่าง ๆ ซึ่งภาระหน้าที่ของสถาบันการศึกษาไม่ใช่เพียงจัดการศึกษาเท่านั้น แต่ยังต้องพัฒนาคุณภาพการให้บริการด้านอื่น ๆ เพื่อให้เป็นสถาบันที่มีคุณภาพสร้างความประทับใจ ความพึงพอใจแก่ผู้เรียนหรือผู้มารับบริการด้วย จึงต้องศึกษาความพึงพอใจของนักศึกษา

เฟซบุ๊กเพจนั้นสามารถเปรียบได้กับภาพลักษณ์ขององค์กรหรือตราสินค้าบนเฟซบุ๊กที่สามารถสร้างได้ฟรี ไม่เสียค่าใช้จ่าย เนื่องจากแฟนเพจสามารถที่จะทำหน้าที่ทางการสื่อสารกับผู้บริโภคผ่านหน้า News Feeds ได้ ด้วยการแทรกซึมเข้าไปในพื้นที่ส่วนตัวของผู้บริโภคที่ใช้งานสื่อเฟซบุ๊ก โดยที่ผู้บริโภคยินยอม นอกจากนี้แฟนเพจยังสามารถให้ข้อมูลได้อย่างต่อเนื่อง มีประสิทธิภาพ และสามารถสื่อสารได้เรียลไทม์ (Real-time) อีกด้วย จึงทำให้แฟนเพจสามารถ

สื่อสารข้อมูลข่าวสารต่าง ๆ ขององค์กรได้อย่างสดใหม่ และทันต่อเหตุการณ์มากกว่าสื่อประเภทอื่น และยังสามารถสื่อสารได้อย่างรวดเร็ว ส่งตรงไปยังกลุ่มเป้าหมายได้มากกว่าด้วย เฟซบุ๊กเพจยังสามารถทำการวัดผลได้จากจำนวนผู้ที่เข้ามากดชื่นชอบ (Like) และการแสดงความคิดเห็น (Comment) ที่หน้าเฟซบุ๊กเพจอันเป็นสิ่งที่บอกได้ถึงประสิทธิภาพในการประชาสัมพันธ์หน้าเฟซบุ๊กเพจต่าง ๆ นั้นได้ว่ามีประสิทธิภาพมากแค่ไหน ผู้วิจัยจึงต้องการศึกษาความรู้สึกของนักศึกษาที่มีต่อมหาวิทยาลัยบนเฟซบุ๊ก เพื่อมหาวิทยาลัยจะได้นำไปใช้เป็นแนวทางในการปรับปรุงและพัฒนาการจัดการเรียนการสอน และพัฒนาระบบการบริการของมหาวิทยาลัยต่อไป

2. การวิเคราะห์และออกแบบขั้นตอนการทำงานของระบบ

กระบวนการการทำงานของระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ (Real-time) ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ (Naive Bayes Classifier) สำหรับภาษาไทย สามารถอธิบายรายละเอียดได้ดังภาพที่ 3.1

ภาพที่ 3.1 กระบวนการทำงานโดยรวมของระบบ

จากภาพที่ 3.1 กระบวนการทำงานของระบบ โดยการรวบรวมข้อมูลจากเฟซบุ๊ก (Facebook) มาเก็บไว้ที่ฐานข้อมูลของระบบ แล้วนำข้อมูลเข้าสู่กระบวนการการตัดคำภาษาไทยโดยใช้วิธีการเทียบคำที่ยาวที่สุด (Longest Matching) โดยจะพิจารณาโดยการวนลูปจากซ้ายไปขวาตามหลักภาษาไทยเพื่อเทียบกับพจนานุกรม Lexicon – Thai แล้วนำมาวิเคราะห์ความเห็นออกมาเป็นคะแนนบวก (+1) ลบ (-1) หรือเป็นกลาง (0) เพื่อสรุปภาพรวมความรู้สึก (Sentiment Analysis) แบบ

2 กลุ่ม คือ ความรู้สึกเชิงบวก (Positive) และความรู้สึกเชิงลบ (Negative) เพื่อจัดเก็บไว้ในฐานข้อมูล และนำผลลัพธ์ที่ได้แสดงผลผ่านหน้าจอของระบบ

2.1 กระบวนการทำงานของระบบในการรวบรวมข้อมูล โดยพนิดา ทรงรัมย์ ได้ศึกษาการจำแนกความคิดเห็นที่มีต่อรัฐประหารในประเทศไทยจากความคิดเห็นบนเฟซบุ๊ก โดยวิธีการรวบรวมข้อมูลด้วย Facebook Graph API (พนิดา ทรงรัมย์, 2559) ซึ่งเป็นวิธีการเรียกใช้ข้อมูลจากเฟซบุ๊กบนอินเทอร์เน็ต (Internet) และเป็นวิธีการเดียวในการเรียกใช้ข้อมูลด้วยระบบคอมพิวเตอร์ งานวิจัยชิ้นนี้จึงใช้วิธีดังกล่าวในการรวบรวมข้อมูล โดยเริ่มจากร้องขอข้อมูลไปยังเฟซบุ๊ก จากนั้นเฟซบุ๊กจะทำการตรวจสอบว่ามีสิทธิ์ใช้งานข้อมูลหรือไม่ หากผ่านการตรวจสอบ เฟซบุ๊กจะส่งข้อมูลกลับมาที่เซิร์ฟเวอร์ (อังกฤษ: server) หรือ เครื่องบริการ หรือ เครื่องแม่ข่าย ด้วยรูปแบบ JSON (JavaScript Object Notation) ระบบจะทำการอ่านและแปลงข้อมูลเพื่อบันทึกลงในฐานข้อมูล ดังภาพที่ 3.2

ภาพที่ 3.2 กระบวนการทำงานของระบบในการรวบรวมข้อมูล

จากภาพที่ 3.2 แสดงกระบวนการทำงานของระบบในการรวบข้อมูล กระบวนการเริ่มจาก การร้องขอข้อมูลไปยังเฟซบุ๊ก จากนั้นจะเข้าสู่กระบวนการที่ 2 คือ การตรวจสอบข้อมูลของเฟซบุ๊กว่ามีสิทธิ์ใช้ข้อมูลหรือไม่ หากผ่านการตรวจสอบเฟซบุ๊กจะส่งข้อมูลกลับมาให้ระบบทำการอ่านข้อมูลและสิ้นสุดในกระบวนการบันทึกข้อมูลสามารถอธิบายรายละเอียดได้ดังต่อไปนี้

2.1.1 กระบวนการร้องขอข้อมูล ในการเรียกใช้ข้อมูลจากทางเฟซบุ๊กนั้นจำเป็นจะต้องมีสิทธิ์ในการใช้ข้อมูล ซึ่งโดยปรกติทั่วไปในการเข้าใช้งานข้อมูล ระบบจะใช้ User และ Password เพื่อยืนยันตัวตนในการเข้าถึงข้อมูล แต่มันก็มีความยุ่งยากอยู่หนึ่งอย่างคือ จะต้องกรอกข้อมูล User และ Password ทุกครั้งที่ต้องการเรียกเข้าดูข้อมูล อีกอย่างคือเรื่องของความปลอดภัย ถ้าหากมีผู้ไม่ประสงค์ดี ดักจับ (Sniffer) การรับส่งข้อมูลระหว่างทาง เมื่อผู้ไม่ประสงค์ดีดักจับ (Sniffer) เอา User และ Password ได้แล้วยังสามารถนำไปใช้ในทางที่ไม่ดีได้ ดังนั้นจึงจำเป็นจะต้องใช้ข้อมูล Access token ในการเข้าถึงข้อมูลทุกอย่าง ทั้งนี้ก็เพื่อความปลอดภัยและความสะดวกในการเรียกใช้งานข้อมูลที่ไม่จำเป็นต้องส่ง User และ Password มาด้วยทุกครั้ง ในกระบวนการขอใช้ข้อมูลจากเฟซบุ๊กนั้นจำเป็นจะต้องแนบ พารามิเตอร์ (Parameter) ชื่อว่า access_token ไปด้วยทุกครั้งดังตัวอย่างภาพที่ 3.1

ตารางที่ 3.1 ตัวอย่างข้อมูลที่ใช้ในการเรียกใช้งาน Facebook Graph API

ชื่อ	**ค่า**	**คำอธิบาย**
HOST	https://graph.facebook.com	Url ของ Facebook API
Page	STOUniversity	
Version	v2.10	เวอร์ชั่นของ Facebook API
Fields	tagged	คอร์ลัมที่ต้องการเลือก
Access token	EAAEMbMqLFZA4BALIPuqLDfh4Ifvaev6 BoTRfDaREKbU2qnK2aWniZC05B35UadI2 kLOZAMWRbINchAchEiTui6cnJwdoWLtiI wogOLoJs6zGNDfcm7UwtwDcVAJDyZA3	Access token ที่ใช้ในการร้องข้อมูล
Comments	comments	ข้อมูลความคิดเห็น
Limit	100	จำนวนที่ต้องการต่อครั้ง
Id	id	รหัสความคิดเห็น
Message	message	ข้อความ
Message tags	Message_tags	Tag ในข้อความ
Created time	created_time	วันที่แสดงความคิดเห็น

จากตารางที่ 3.1 สามารถนำข้อมูลมาประกอบเป็น URLในรูปแบบดังภาพที่ 3.3

```
https://graph.facebook.com/v2.10/{page}?fields=tagged{comments.limit(100){id,message,message_tags,created_time},link}&access_token={access_token}
```

ภาพที่ 3.3 ตัวอย่าง URL และชุดข้อมูลที่ใช้ในการเรียกใช้ข้อมูล Facebook Graph API

2.1.2 กระบวนการตรวจสอบข้อมูล ในกระบวนการเฟซบุ๊กจะเป็นฝ่ายตรวจสอบความถูกต้องของข้อมูลโดยถ้าหากผ่านการตรวจสอบข้อมูลชุดพารามิเตอร์ (Paramiter) ที่ส่งไปพร้อมกับผ่านการตรวจสอบสิทธิ์ในการใช้งานข้อมูล Facebook Graph API จะส่งข้อมูลกลับมาแต่หากไม่ผ่าน Facebook ก็จะตอบกลับมาเช่นกันพร้อมกับระบุว่าเกิดจากสาเหตุใดดังตัวอย่างตามภาพที่ 3.4

```
{
  "error": {
    "message": "(#10) To use 'Page Public Content Access', your use of this endpoint must be reviewed and approved by Facebook. To submit this 'Page Public Content Access' feature for review please read our documentation on reviewable features: https://developers.facebook.com/docs/apps/review.",
    "type": "OAuthException",
    "code": 10,
    "fbtrace_id": "FnvB0gGAFRO"
  }
}
```

ภาพที่ 3.4 ตัวอย่างข้อมูลการตอบกลับจาก Facebook Graph API กรณีเกิดข้อผิดพลาด

2.1.3 กระบวนการส่งข้อมูล เมื่อคำร้องขอใช้งานผ่านการตรวจสอบระบบจะส่งค่ากลับมายังเซิฟเวอร์ของระบบในรูปแบบ JSON ซึ่งจะมีลักษณะดังภาพที่ 3.5

```
- comments: {
    - data: [
        - {
            created_time: "2017-12-15 08:49:59",
          - from: {
                name: "ขุนภูมิ ขุนภูมิ ขุนภูมิ",
                id: "1919479198372587"
            },
            message: "แชร์ได้ไหมครับท่านอาจารย์",
            id: "1687509007936193_1687604971259930"
        },
        - {
            created_time: "2017-12-15 12:24:39",
          - from: {
                name: "Tonmong Maha Kyaw Kyaw",
                id: "1940852459264748"
            },
            message: "นับถือ respect มากค้าบ",
            id: "1687509007936193_1687769117910182"
        },
        - {
            created_time: "2017-12-15 08:06:50",
          - from: {
                name: "Linda Cartlidge",
                id: "1726383717413851"
            },
            message: "Lovely photo liz😘",
            id: "1687509007936193_1687575811262846"
        }
```


ภาพที่ 3.5 ข้อมูลความคิดเห็นจากเฟซบุ๊กในรูปแบบ JSON

2.1.4 กระบวนการการอ่านข้อมูล หลังจากได้ที่ข้อมูลความคิดเห็นมา ระบบจะทำการแปลงข้อมูลโดยใช้ความสามารถของภาษา PHP ที่มีฟังก์ชันถอดรหัสข้อมูลชนิด JSON อยู่แล้ว โดยจะแปลงข้อมูลจาก JSON เป็นชนิดอาร์เรย์ (Array) ดังตัวอย่างโค้ดการถอดรหัสข้อมูลจาก JSON เป็นอาร์เรย์ดังภาพที่ 3.6

```
$dataJsonDecode = json_decode($jsonData, true);
```

ภาพที่ 3.6 ตัวอย่างโค้ดการถอดรหัสข้อมูลจาก JSON เป็นข้อมูลชนิดอาร์เรย์

2.1.5 กระบวนการบันทึกข้อมูล ในกระบวนการนี้หลังจากที่ได้ชุดข้อมูลอาร์เรย์มาเรียบร้อยแล้ว ระบบจะทำการจัดรูปแบบข้อมูล เพื่อให้ Field ของข้อมูลตรงกันกับชุดข้อมูลในฐานข้อมูล ด้วยการวนลูป (Foreach) จากนั้นระบบจะทำการบันทึกลงฐานข้อมูลความคิดเห็น

2.2. กระบวนการตัดคำภาษาไทย ใช้การจับคู่ที่ยาวที่สุด (Longest Matching) โดยจะพิจารณาโดยการวนลูปจากซ้ายไปขวาตามหลักภาษาไทยเพื่อเทียบกับพจนานุกรม Lexicon – Thai สามารถอธิบายด้วยอัลกอริทึม (Algorithm) การตัดคำภาษาไทยโดยมีขั้นตอนดังต่อไปนี้

ขั้นตอนที่ 1 แยกอักขระ เมื่อนำข้อความเข้าสู่กระบวนการตัดคำสิ่งแรกที่ทำคือแยกอักขระทุกตัวออกจากกันแล้วจัดเก็บเข้าไปในตัวแปรชนิดอาร์เรย์จากนั้นโปรแกรมจะนับว่ามีอักขระทั้งหมดกี่ตัวเพื่อหาความยาว แต่ตัวแปรชนิดอาร์เรย์เริ่มนับตำแหน่งจากศูนย์ จึงจำเป็นต้องลบด้วย 1 เพื่อบอกตำแหน่งที่ถูกต้องทั้งหมดของอักขระ

ขั้นตอนที่ 2 ค้นหาคำศัพท์ เริ่มทำงานซ้ำจนกว่าค่าของตัวแปรระบุตำแหน่ง จะมีค่ามากกว่าตัวแปรความยาวของอักขระ โดยแต่ละรอบจะนำข้อความไปเทียบกับพจนานุกรมหากพบข้อความจะเก็บไว้ในตัวแปรคำที่พบในข้อความพร้อมกับระบุตำแหน่งที่พบ หากไม่พบจะนำอักขระต่อกับอักขระก่อนหน้าแล้วทำการเทียบกับพจนานุกรมอีกครั้ง

ขั้นตอนที่ 3 ค้นหาคำศัพท์อื่น ๆ กรณีที่พบคำศัพท์ในอักขระ โปรแกรมจะทำงานซ้ำ (For loop) โดยเริ่มจากตำแหน่งของอักขระปัจจุบันไปจนถึงตำแหน่งอักขระสุดท้าย โดยแต่ละรอบจะนำข้อความไปเทียบกับพจนานุกรมหากพบข้อความจะเก็บไว้ในตัวแปรคำที่พบในข้อความพร้อมกับระบุตำแหน่งที่พบ เพื่อแยกคำที่พบออกจากกัน

ขั้นตอนที่ 4 เลือกคำที่ยาวที่สุด ถ้าโปรแกรมพบว่าตัวแปรคำที่พบในข้อความไม่ใช่ค่าว่าง โปรแกรมจะเลือกคำศัพท์ที่ยาวที่สุดออกมา ด้วยหลักการอย่างง่ายคือเลือกตำแหน่งของอาร์เรย์ในลำดับสุดท้ายที่พบคำศัพท์ออกมา จะได้ข้อความที่ยาวที่สุดในแต่ละรอบแล้วเก็บข้อความที่ได้ไว้ในตัวแปรผลลัพธ์การตัดคำ พร้อมทั้งตั้งค่าตัวแปรคำที่พบและตัวแปรข้อความที่พบชนิดอาร์เรย์ เป็นค่าเริ่มต้น หลังจากนั้นโปรแกรมจะทำซ้ำ (While Loop) ข้อความในลำดับถัดไปโดยเริ่มจากตำแหน่งที่พบข้อความที่มากที่สุดจนกว่าค่าระบุตำแหน่ง (Pointer) จะมีค่ามากกว่าความยาวของอักขระ

ขั้นตอนที่ 5 ตรวจสอบและส่งออก ในกรณีที่ไม่พบคำศัพท์ใด ๆ เลยโปรแกรมจะนำข้อความที่ไม่สามารถตัดได้ให้เป็นผลลัพธ์ของการตัดคำ แต่หากพบคำศัพท์ข้อความที่ถูกตัดจะอยู่ในรูปแบบของอาร์เรย์เป็นคำที่ถูกตัดเรียบร้อยแล้วส่งออกไป สามารถเขียนเป็นแผนภาพรหัสเทียม (Pseudocode) ได้ดังภาพที่ 3.7

```
Input: Sentences
Output: Word Token
Read Sentences; Strings  = Extract(Sentences)
While(Pointer=0; Pointer <= CountStringLength(Strings) -1)
        TempWord += String[Pointer] + TempWord
        If(TempWord === Dictionary's word)
                FoundWord['word']  = TempWord
                For(pointer +1 > CountStringLength(Strings) -1; Pointer ++ )
                        MoreStringTemp += MoreStringTemp +
                                String[Pointer]
                        If(MoreStringTemp === Dictionary's word)
                                FoundWord['word']  = MoreStringTemp
                ResultToken[] += Max(FoundWord)
        Pointer++
Return ResultToken
```

ภาพที่ 3.7 ขั้นตอนกระบวนการตัดคำออกจากข้อความภาษาไทย

จากภาพที่ 3.7 เป็นกระบวนการตัดคำภาษาไทยซึ่งสอดคล้องกับ 5 ขั้นตอนที่กล่าวไว้ข้างต้น โดยมีข้อมูลที่นำเข้าคือ ข้อความความคิดเห็น และผลลัพธ์ที่ได้คือ ข้อความที่ถูกตัดเรียบร้อยที่อยู่ในรูปแบบของข้อมูลประเภทแถวลำดับ (Array) เพื่อส่งไปขั้นตอนถัดไป

2.3 การจำแนกข้อมูลโดยใช้ตัวจำแนกนาอีฟ เบย์ (Naive Bayes Classifier) โดยการให้น้ำหนักคำในประโยค ด้วยวิธีการเทียบคำที่ถูกตัดกับฐานข้อมูลประเภทของคำ (Daniel Jurafsky and James H. Martin, 2018) ซึ่งมี 5 ประเภทด้วยกัน ในแต่ละประเภทของคำมีประเภทที่สนใจนำมาคิดค่าคะแนน 3 ประเภทคือคำที่สถานะเป็นกลาง, คำที่มีสถานะเชิงลบ และคำที่มีสถานะเชิงบวก ดังตารางที่ 3.2

ตารางที่ 3.2 ประเภทของคำในพจนานุกรม

ตัวอย่างคำ	ประเภท	รายละเอียด
นางสาว	Prefix	คำนำหน้าชื่อ
และแล้ว	Ignore	คำที่ไม่สนใจ
สังคม	Neutral	คำที่มีสถานะเป็นกลาง
ดีมาก ๆ	Positive	คำที่มีสถานะเป็นเชิงบวก
เลวทราม	Negative	คำที่มีสถานะเป็นเชิงลบ

เมื่อมีการตัดคำและกำหนดประเภทของคำภายในพจนานุกรมแล้วจะมีการจำแนกข้อมูลโดยการให้ค่าน้ำหนักของคำภายในประโยค เพื่อนำมาวิเคราะห์ความเห็นออกมาเป็นคะแนนเพื่อวิเคราะห์ความรู้สึก (นริศร์ พรหมบุตร, 2550).(นภดล สิทธิเลิศ, 2556) แบบ 3 กลุ่ม คือ ความรู้สึกเชิงบวก (Positive) ความรู้สึกเชิงลบ (Negative) และเป็นกลาง สามารถอธิบายด้วยอัลกอริทึม (Algorithm) การวิเคราะห์ความรู้สึก ด้วยทฤษฎีนาอีฟ เบย์ (Naive Bayes Classifier) โดยมีขั้นตอนแยกย่อยดังนี้

ขั้นตอนที่ 1 ตั้งค่าตัวแปร โดยแต่ละตัวแปรจะมีค่าดังนี้ ตัวแปร Classes มีค่าด้วยกัน 3 ค่าคือ Positive, Negative, Neutral ตัวแปรค่าความน่าจะเป็นของแต่ละประเภทความรู้สึก (Prior) มีค่าเท่ากับ 0.3333 เท่ากัน เพราะความน่าจะเป็นของแต่ละเหตุการณ์มีค่าเท่ากับ 1/3 เท่ากับ 0.3333 และค่าคะแนนของแต่ละประเภทความรู้สึกมีค่าเท่ากับ 1

ขั้นตอนที่ 2 ทำซ้ำข้อความที่ถูกตัด นำคำที่แยกออกมาจากข้อความไปเทียบกับคำในตารางประเภทของคำ หากพบ ให้นำคำที่พบไปตรวจสอบว่าอยู่ในข้อความประเภทใด แล้วบวกค่าคะแนนของประเภทความรู้สึกนั้นด้วย 1 และเก็บไว้ที่ตัวแปรค่าคะแนนของแต่ละประเภทความรู้สึก (Score) ของคลาสนั้น ๆ ทำจนกว่าจะครบทุกคำในประโยค หลังจากนั้นนำค่าคะแนนของแต่ละคลาสไปคูณกับค่าความน่าจะเป็นของแต่ละประเภทความรู้สึก (Prior) และเก็บไว้ที่ตัวแปร (Score)

ขั้นตอนที่ 3 หาค่าคะแนนผลรวม ค่าความน่าจะเป็นของข้อความมีโอกาสเกิดขึ้นได้เพียงแค่ 1 แต่ในขั้นตอนที่ 2 นั้น หากพบข้อความจำนวนมาก จะทำให้ค่าคะแนนของแต่ละประเภทความรู้สึก (Score) มีค่ามากกว่า 1 ดังนั้นจึงจำเป็นต้องนำผลรวมของค่าความรู้สึกแต่ละประเภทไปเป็นตัวหาร เพื่อให้ได้ค่าความน่าจะเป็นที่ถูกต้อง โดยการนำค่าที่ได้ของแต่ละประเภทความรู้สึกมาหาผลรวม ด้วยการทำซ้ำ (Foreach) ค่าคะแนนของแต่ละประเภทความรู้สึก (Score) ใน

แต่ละรอบค่าผลรวมของค่าคะแนนของแต่ละประเภทความรู้สึกจะถูกบวกเพิ่มด้วยค่าคะแนนของแต่ละประเภทความรู้สึกของคลาสนั้น จนกว่าจะครบทุกคลาส

ขั้นตอนที่ 4 หาค่าคะแนนความรู้สึก เป็นการหาค่าความน่าจะเป็นที่แท้จริงของโอกาสที่จะเกิดขึ้นของแต่ละประเภทความรู้สึก ด้วยการนำค่าคะแนนของแต่ละประเภทความรู้สึก (Score) หารด้วยค่าผลรวมของค่าคะแนนของแต่ละประเภทความรู้สึก (TotalScore) แล้วเก็บกลับคืนไว้ที่ ค่าคะแนนของแต่ละประเภทความรู้สึก (Score) จนกว่าจะครบทุกคลาส

ขั้นตอนที่ 5 เรียงลำดับและส่งออก นำค่าคะแนนของแต่ละประเภทความรู้สึก (Score) มาเรียงด้วยฟังก์ชัน SortDesc หลังจากที่เรียงเสร็จแล้ว นำค่าของผลการเรียงลำดับมาหาค่าที่มากที่สุดด้วยฟังก์ชัน Max หากค่าของคลาสใดมีคะแนนมากที่สุด หมายความว่าข้อความนี้มีค่าเป็นคลาสนั้นหรือข้อความนี้มีความรู้สึกตามประเภทที่มีค่ามากสุด และส่งออกผลการหาค่าคะแนนพร้อมระบุว่าเป็นข้อความประเภทใด จากขั้นตอนทั้งหมดที่กล่าวไว้ข้างต้นสามารถเขียนเป็นแผนภาพรหัสเทียม (Pseudocode) ได้ดังภาพที่ 3.8

```
Input: Words
Output: Sentiment Classification
Set Classes = {positive, negative, neutral}
Set Prior = {positive = 0.3333, negative = 0.3333, neutral = 0.3333}
Set Score = 1
Foreach(words as word)
        If (word === dictionary's word)
                Score[Class] += Score[Class]
        Endif
End foreach
Score[Class] = Prior[Class] * Score[Classes]
Foreach (Classes as Class)
                Totalscore += Score[class]
End foreach
Foreach (Classes as class)
                Score[Class] = Score[Class] / Totalscore
End foreach
Classification = Max(SortDesc(Score[Class]))
Return Classification
```

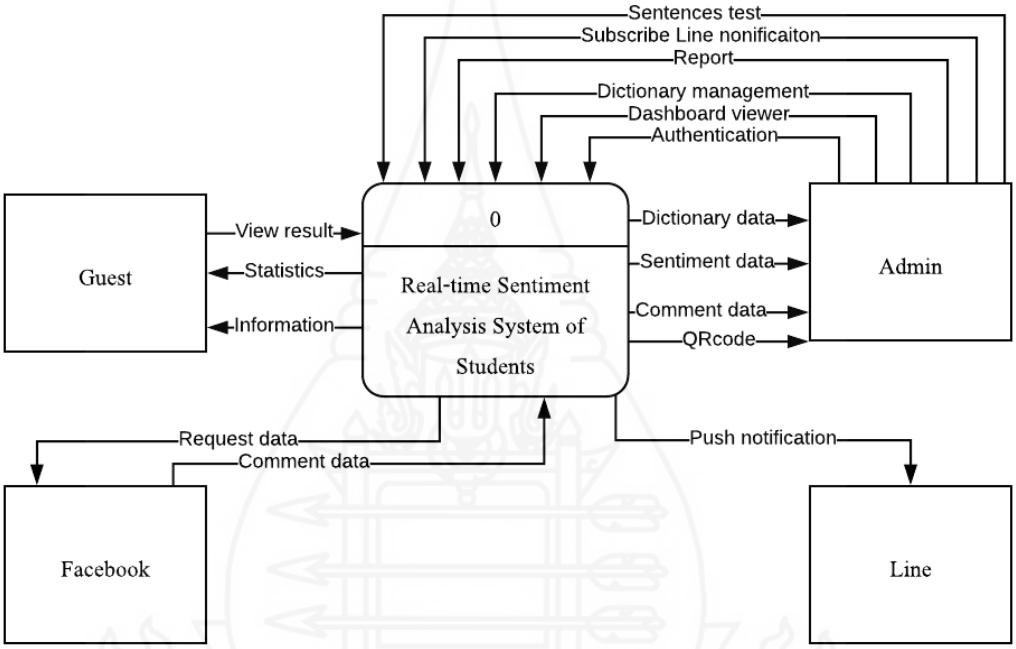
ภาพที่ 3.8 ขั้นตอนกระบวนการวิเคราะห์ความรู้สึกจากข้อความ

จากภาพที่ 3.8 เป็นขั้นตอนกระบวนการวิเคราะห์ความรู้สึกจากข้อความ ซึ่งมีความสอดคล้องกับ 5 ขั้นตอนข้างต้น โดยนำเข้าข้อมูลที่ผ่านกระบวนการตัดคำมาแล้ว มาคำนวณหาค่าคะแนนและส่งออกผลลัพธ์การวิเคราะห์ความรู้สึกว่าข้อความเป็นประเภทใด โดยการคำนวณหาค่าคะแนนของแต่ละคลาสแล้วเลือกคลาสที่มีค่าคะแนนสูงสุด สามารถคำนวณหาค่าคะแนนด้วยสมการที่ 3.1 (Vivek Narayanan, Ishan Arora, and Arjun Bhatia, 2013)

$$P(c_i|d) = \frac{P(d|c_i)\ x\ P(c_i)}{P(d)}$$

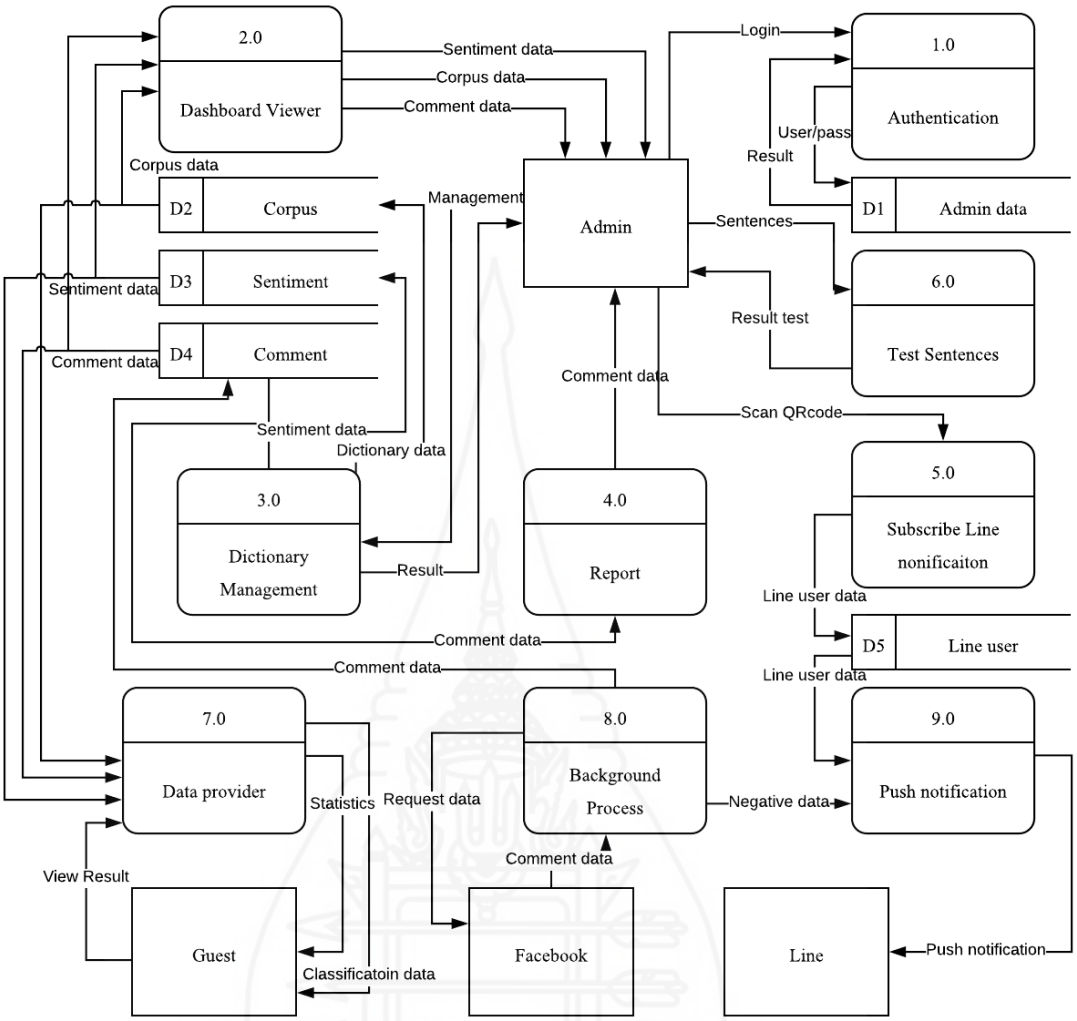
จากสมการที่ 3.1 เป็นการหาค่าคะแนนทั้งหมดของเอกสาร โดยแทน d คือเอกสาร (Document) ในการหาค่าน้ำหนักคะแนนของโอกาสที่เอกสารนี้จะเป็นคลาสใด ๆ หมายความว่าในการคำนวณค่าคะแนนนั้นไม่ได้นับคะแนนแบบแบ่งประโยค แต่เป็นการหาค่าคะแนนทั้งหมดของเอกสารสารในรอบเดียวแล้วมาคำนวณหาค่าคะแนน

2.4 การออกแบบแผนภาพการไหลของข้อมูลระดับบนสุด (Context Diagram) เพื่อแสดงภาพรวมการทำงานของระบบ การไหลของข้อมูล ส่วนที่เกี่ยวข้องทั้งหมดกับระบบดังภาพที่ 3.9



ภาพที่ 3.9 แผนผังงานระบบ (Context Diagram)

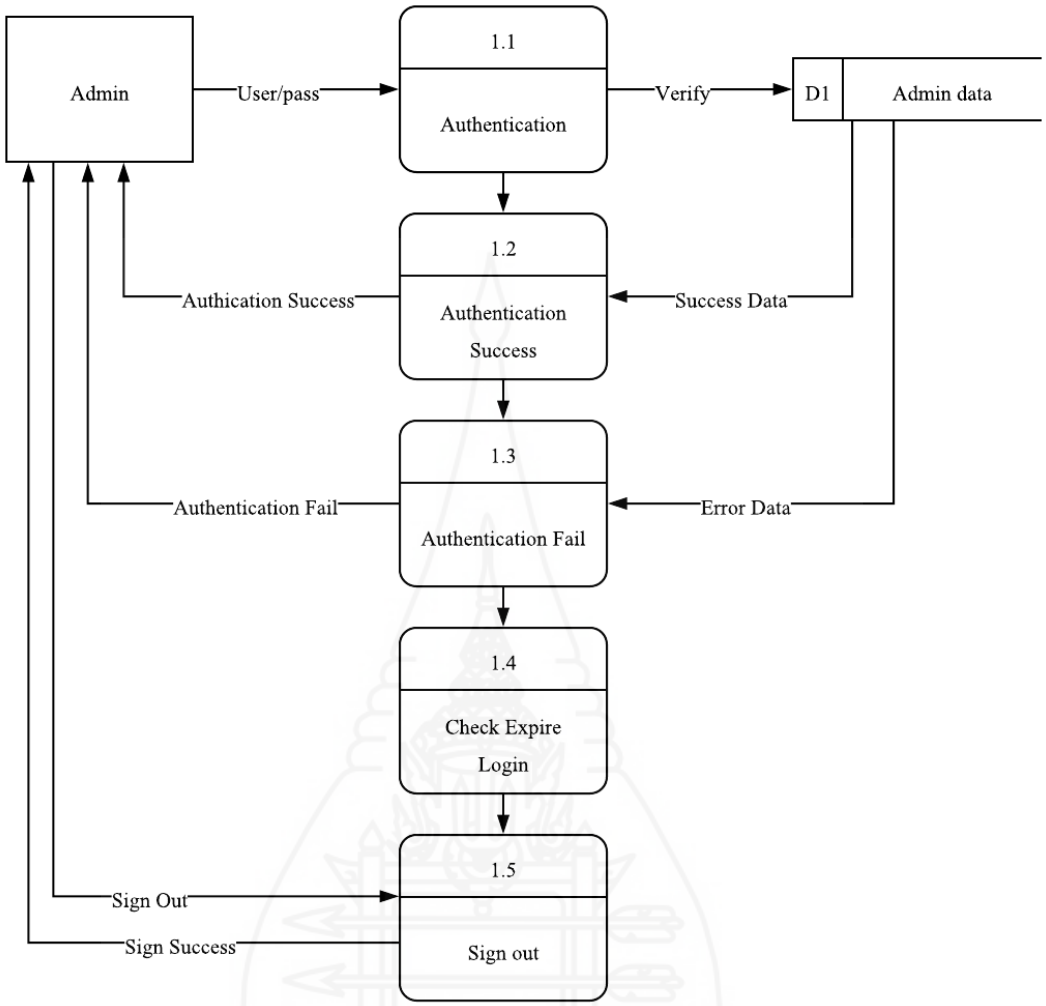
จากภาพที่ 3.9 เป็นการแสดงภาพรวมของระบบโดยมี 4 ส่วนงานที่เกี่ยวข้องกับระบบ คือ ผู้ใช้งานทั่วไป (Guest) ผู้ดูแลระบบ (Amin) Facebook ใช้เป็นแหล่งข้อมูลในการสืบค้นข้อมูลความคิดเห็นของนักศึกษาบนเฟซบุ๊กเพจ มหาวิทยาลัยสุโขทัยธรรมาธิราช และ Line ใช้ในการแจ้งเตือนข้อความไปยังผู้ดูแลระบบสามารถแบ่งแยกย่อยได้ดังแผนภาพที่ 3.10



ภาพที่ 3.10 แผนภาพการไหลของข้อมูลระดับที่ 1 (Dataflow Diagram Level 1)

จากภาพที่ 3.10 จะเห็นได้ว่ามีกระบวนการทำงานแยกย่อยทั้งหมดอยู่ 9 กระบวนการด้วยกัน และมีส่วนที่ใช้งานรวมถึงการให้บริการข้อมูลอยู่ 4 Entity คือ 1) ผู้ใช้งานทั่วไป, 2) ผู้ดูแลระบบ, 3) Facbook และ 4) Line ซึ่งจะอธิบายรายละเอียดการทำงานของส่วนต่าง ๆ ดังต่อไปนี้

2.4.1 การจัดการการเข้าถึงข้อมูล (Authentication) ในส่วนของกระบวนการนี้จะเป็นการออกแบบระบบเพื่อให้มีความปลอดภัยในการเข้าถึงข้อมูลในระบบ โดยจะมีการตรวจสอบข้อมูลทุกครั้งในการเข้าถึงระบบบริการจัดการข้อมูล (Backend) หากไม่ผ่านการตรวจสอบ ก็จะไม่สามารถเข้าถึงข้อมูลได้ รวมไปถึงกระบวนการออกจากระบบกรณีที่ไม่ต้องการจะใช้งานระบบต่อเพื่อป้องกันไม่ให้บุคคลอื่นสามารถเข้ามาใช้งานระบบได้ สามารถอธิบายรายละเอียดการไหลข้อมูลดังภาพที่ 3.11ดังนี้

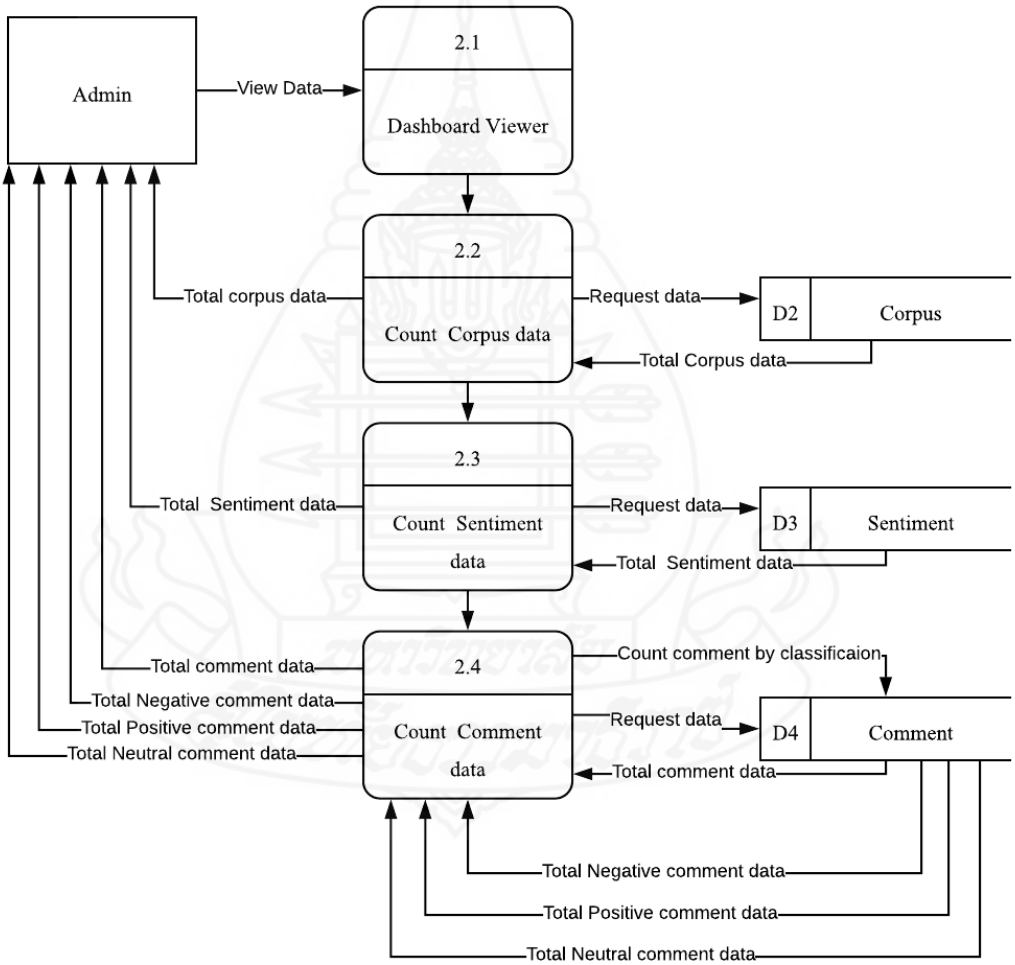


ภาพที่ 3.11 กระบวนการจัดการการเข้าถึงข้อมูล (Authentication)

จากภาพที่ 3.11 สามารถอธิบายรายละเอียดได้ดังนี้ โดยเริ่มจากเมื่อผู้ดูแลระบบกรอกข้อมูลบัญชีผู้ใช้ (User) และรหัสผ่าน (Password) ส่งเข้ามาในระบบ เพื่อให้ระบบทำการตรวจสอบ จากนั้นกระบวนการยืนยันตัวตน (Authentication) จะทำการนำข้อมูลไปตรวจสอบกับฐานข้อมูลผู้ดูแลระบบ (Admin Data) สิ่งที่เกิดขึ้นจะมี 2 กรณีด้วยกันคือ 1) เมื่อผ่านการตรวจสอบ ข้อมูลจะถูกส่งกลับมายังกระบวนการตรวจสอบสำเร็จ (Authentication Success) แล้วส่งผลลัพธ์การเข้าสู่ระบบสำเร็จยังผู้ใช้ 2) ไม่ผ่านการตรวจสอบข้อมูลจะถูกส่งมายังกระบวนการไม่ผ่านการยืนยันตัวตน (Authentication Fail) จากนั้นก็จะส่งข้อมูลไปให้ผู้ดูแลระบบ กระบวนการที่ 3 กระบวนการตรวจสอบเวลาหมดอายุที่อยู่ในระบบ (Check Expire Login) เป็นกระบวนการคอยตรวจสอบอายุการเข้าใช้งานของผู้ดูแลระบบซึ่งผู้ดูแลระบบจะสามารถอยู่ในระบบได้เป็นเวลา 24 ชั่วโมงทั้งนี้ก็

เพื่อกรณีที่ผู้ดูแลระบบลืมออกจากระบบ เมื่อครบระยะเวลาที่กำหนดระบบจะตรวจสอบและทำการออกจากระบบให้อัตโนมัติ นอกจากนี้ยังมีกระบวนการออกจากระบบ (Sign out) เมื่อผู้ดูแลระบบไม่ต้องการที่จะใช้งานระบบต่อ และเพื่อป้องกันการเข้ามาสวมรอยใช้งาน โดยผู้ดูแลระบบจะส่งข้อมูลการขอออกจากระบบมาที่กระบวนการออกจากระบบ (Sign out) ระบบจะทำการให้ผู้ดูแลระบบออกจากระบบ และส่งข้อมูลกลับไปยังผู้ดูแลระบบเป็นอันเสร็จสิ้นกระบวนการ

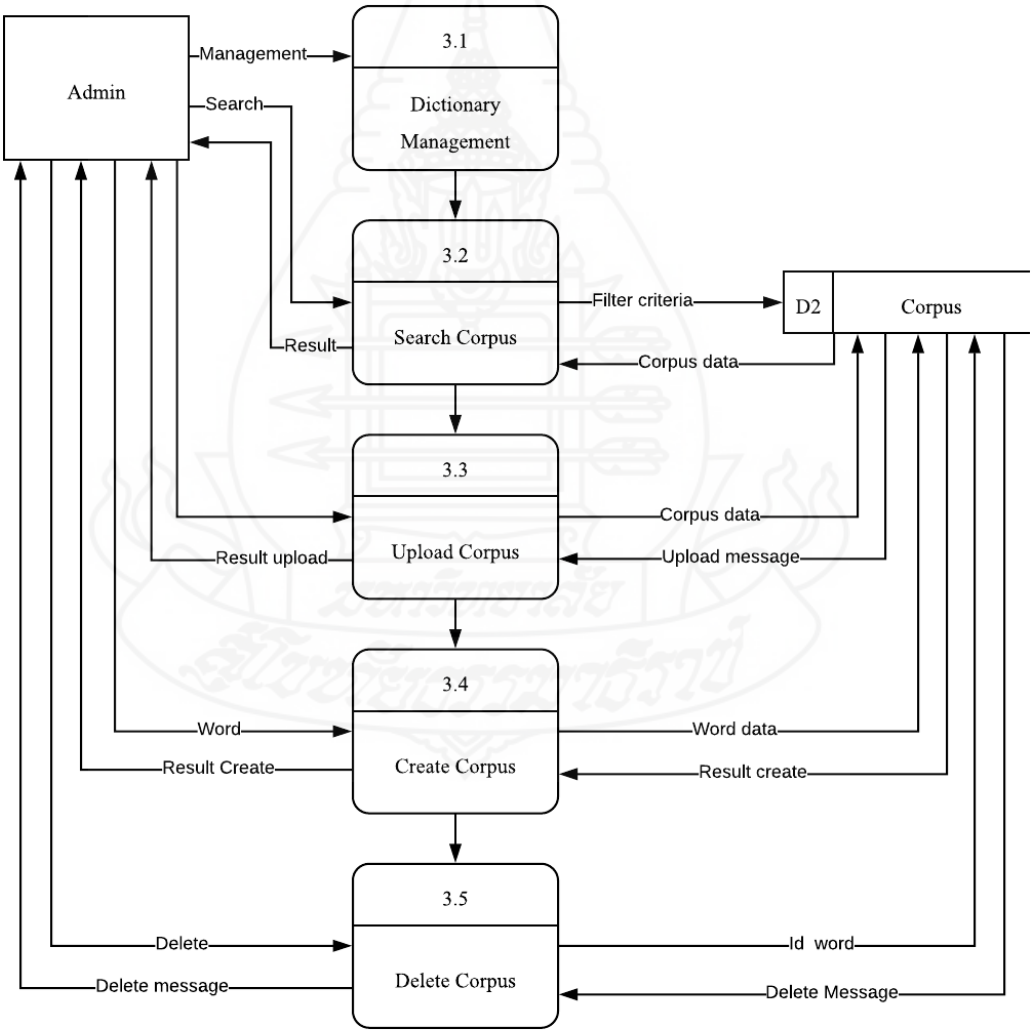
2.4.2 การดูรายงานสถิติต่าง ๆ (Dashboard Viewer) ในส่วนของกระบวนการนี้จะเป็นการนำเสนอข้อมูลต่าง ๆ ให้กับผู้ดูแลระบบให้ทราบถึงตัวเลขข้อมูลโดยรวมที่มีอยู่ในระบบ เช่น จำนวนคลังคำศัพท์ ณ ปัจจุบัน จำนวนฐานข้อมูลคลังคำศัพท์ที่แยกตามประเภทความรู้สึกและข้อมูลอื่น ๆ สามารถเขียนอธิบายการไหลของข้อมูลได้ดังภาพที่ 3.12 ดังนี้



ภาพที่ 3.12 กระบวนดูรายงานสถิติต่าง ๆ (Dashboard Viewer)

จากภาพที่ 3.12 จะเห็นได้ว่ามีกระบวนการแยกย่อยของการดูรายงานสถิติอยู่ด้วยกัน 4 กระบวนการโดยเริ่มจาก กระบวนการที่ 2.1) เมื่อผู้ดูแลระบบส่งคำร้องในการขอดูข้อมูลรายงานข้อมูลเข้ามา จากนั้นนั้นระบบจะส่งคำร้องต่อไปยังกระบวนการ 2.2) คือกระบวนการนับจำนวนข้อมูลของฐานข้อมูลคลังคำศัพท์โดยใช้ข้อมูล ณ ขณะนั้นของฐานข้อมูลคลังคำศัพท์ (Corpus) หลังจากนั้นจะส่งข้อมูลไปยังกระบวนการที่ 2.3) เพื่อนับจำนวนคำในฐานข้อมูลแยกตามประเภทความรู้สึก (Sentiment Data) และกระบวนการสุดท้ายคือกระบวนการ 2.4) คือนับจำนวนความคิดเห็นทั้งหมด เมื่อเสร็จทุกกระบวนการ ผู้ดูแลระบบจะได้ข้อมูลสถิติต่าง ๆ ที่มีอยู่ในระบบ

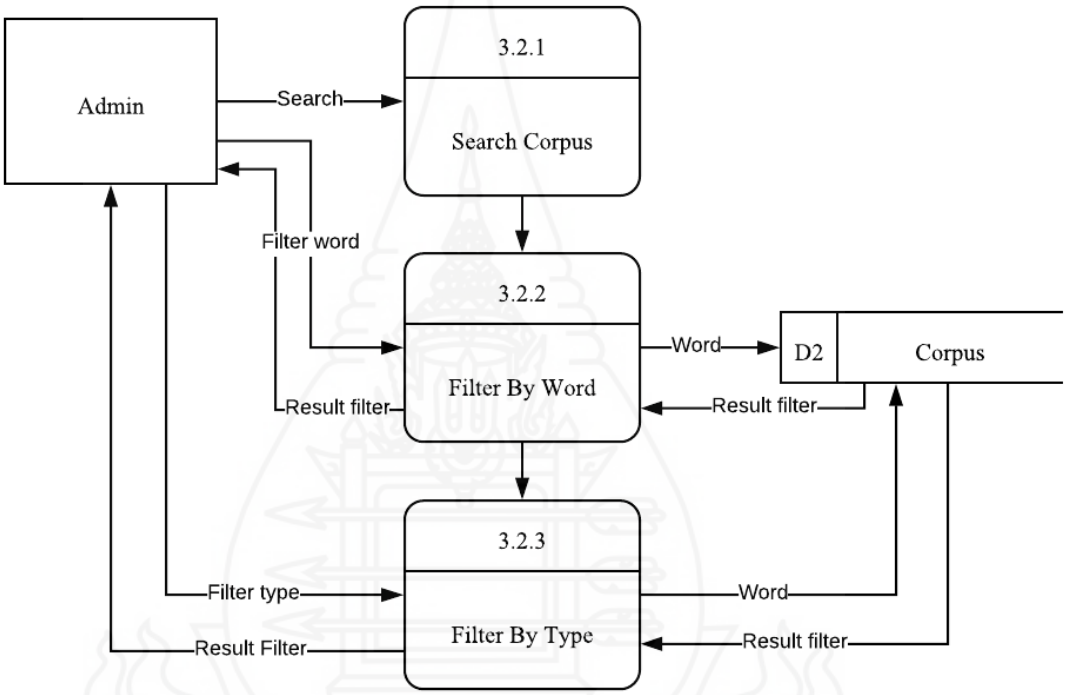
2.4.3 กระบวนการจัดการคลังศัพท์ที่อยู่ในระบบ (Dictionary Management) ในส่วนนี้ผู้ดูระบบสามารถ บริหารจัดการโดยสามารถเพิ่ม ลบ แก้ไข และค้นหาข้อมูลคลังคำศัพท์ผ่านส่วนระบบงานนี้ได้ สามารถอธิบายการไหลของข้อมูลได้ดังภาพที่ 3.13 ดังต่อไปนี้



ภาพที่ 3.13 กระบวนการจัดการข้อมูลคลังคำศัพท์ (Dictionary Management)

จากภาพที่ 3.13 จะเห็นได้ว่ามีกระบวนการแยกย่อยอยู่ 5 กระบวนการด้วยกัน ซึ่งกระบวนการเริ่มขึ้นจากเมื่อผู้ดูแลระบบต้องการจะดำเนินการแก้ไขข้อมูลคลังคำศัพท์ที่มีอยู่ในระบบ เช่นการค้นหาข้อมูล (Search Corpus) การอัพโหลดคลังคำศัพท์ (Upload Dictionary) การเพิ่มคำศัพท์ใหม่ (Create Corpus) และการลบข้อมูลคำศัพท์ออกจากระบบ (Delete Corpus)

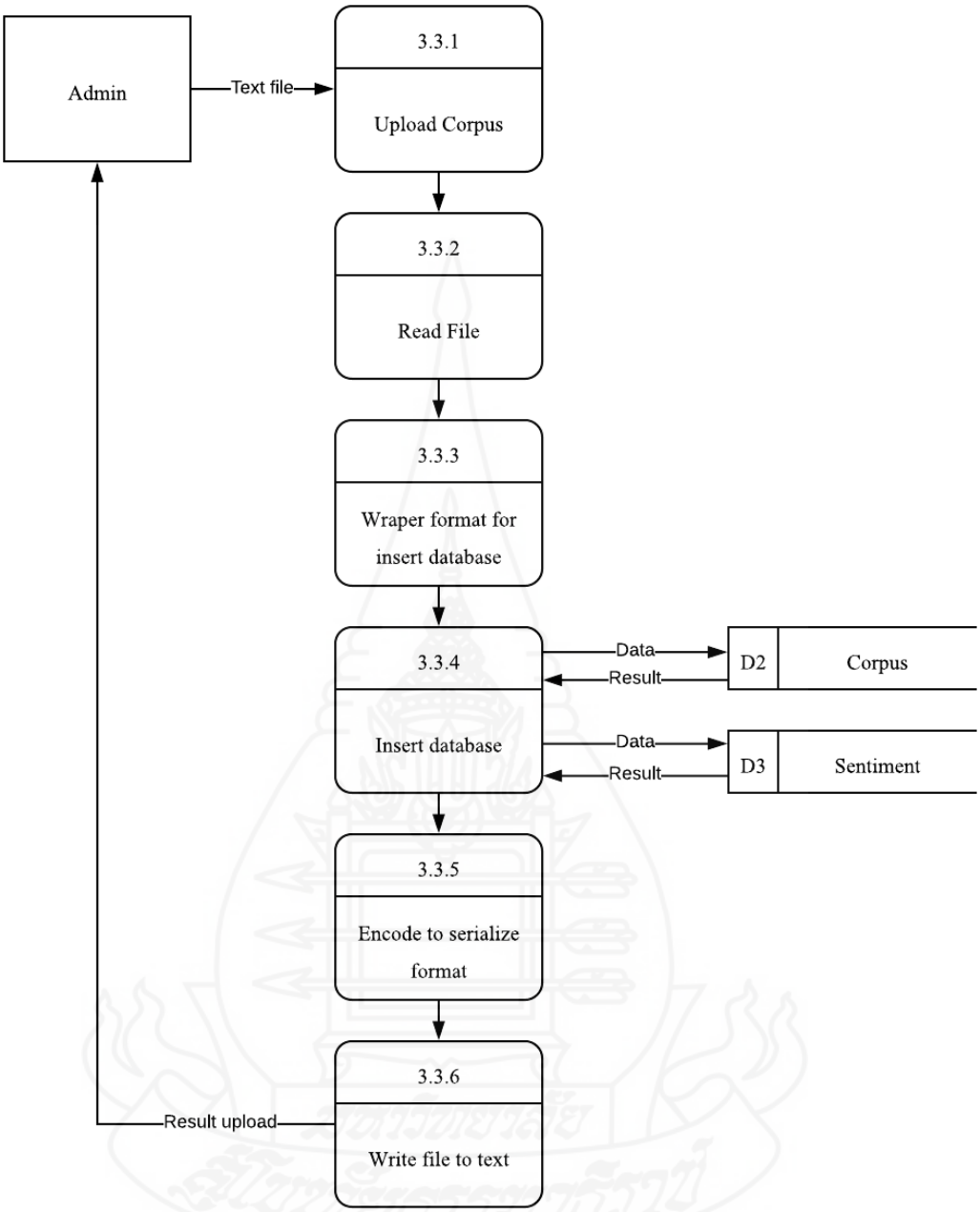
1) กระบวนการค้นหาคลังคำศัพท์ (Search Dictionary) เป็นกระบวนการที่ออกแบบมาให้ผู้ดูแลระบบสามารถค้นหาข้อมูลคลังคำศัพท์ได้ รวมถึงการค้นหาตามประเภทของคำศัพท์ ซึ่งสามารถอธิบายขั้นตอนการไหลของข้อมูลได้ดังภาพที่ 3.14 ดังนี้



ภาพที่ 3.14 กระบวนการค้นหาคลังคำศัพท์ (Search Corpus)

จากภาพที่ 3.14 เป็นกระบวนการสืบค้นข้อมูลในฐานข้อมูลคลังคำศัพท์ ซึ่งจะประกอบไปด้วย กระบวนการแยกย่อยอยู่ 3 ขั้นตอนด้วยกันคือ 3.2.1) กระบวนการสืบคืนข้อมูลทั้งหมด 3.2.2) กระบวนการสืบค้นข้อมูลแบบแยกตามคำศัพท์ และ3.2.3) กระบวนการสืบคืนข้อมูลตามประเภทของคำศัพท์ ทั้งนี้กระบวนการสืบค้นข้อมูลได้ใช้แหล่งข้อมูลจาก ฐานข้อมูลคลังศัพท์ (Corpus) และข้อมูลที่ผู้ดูแลระบบจะได้รับคือผลลัพธ์การสืบค้นข้อมูลตามเงื่อนไขของผู้ดูแลระบบที่เลือกเข้ามา

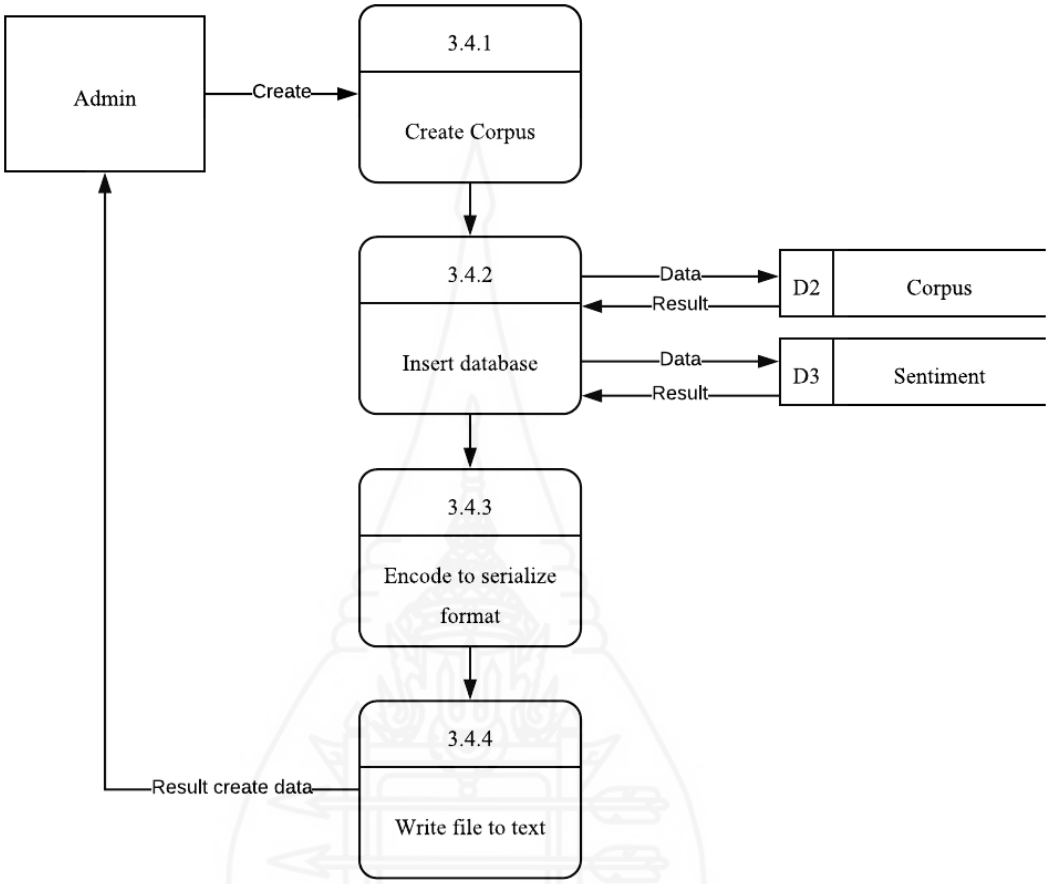
2) กระบวนการอัพโหลดคลังคำศัพท์ (Upload Corpus) สำหรับกระบวนการนี้เป็นส่วนหนึ่งของกระบวนการบริหารจัดการคลังคำศัพท์ มีรายละเอียดดังภาพที่ 3.15



ภาพที่ 3.15 กระบวนการอัพโหลดคำศัพท์ (Upload Corpus)

จากภาพที่ 3.15 เป็นกระบวนการอัพโหลดคลังคำศัพท์โดยประกอบไปด้วยกระบวนการแยกย่อยทั้งหมด 6 คือ 3.3.1) กระบวนการอัพโหลด .3.3.2) กระบวนการอ่านไฟล์ 3.3.3) กระบวนการเตรียมข้อมูล 3.3.4) กระบวนการเพิ่มข้อมูลลงในฐานมูลทั้งสองส่วน 3.3.5) กระบวนการเข้ารหัสข้อมูล และ 3.3.6) กระบวนการเขียนไฟล์กลับไปยัง Text File เพื่อนำไปใช้งานในกระบวนการตัดคำและวิเคราะห์ความรู้สึก

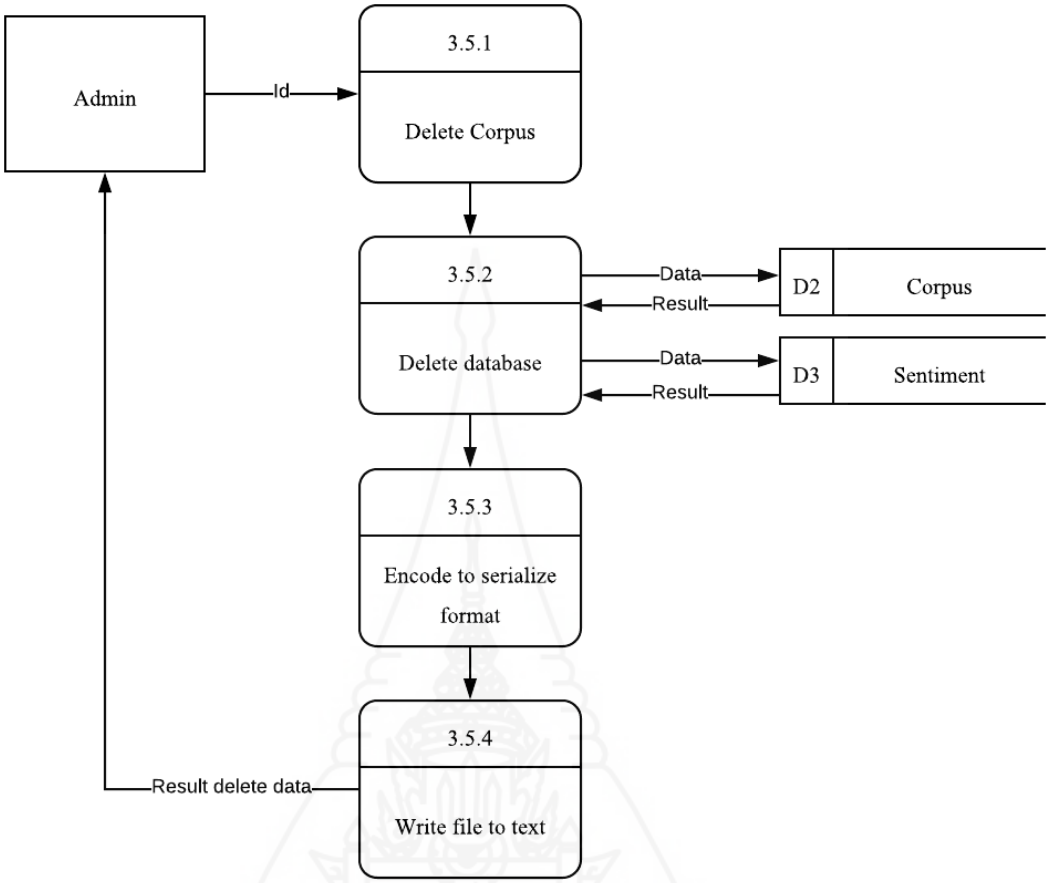
3) กระบวนการเพิ่มคำศัพท์ใหม่ (Create Corpus) สำหรับกระบวนการนี้เป็นส่วนหนึ่งของกระบวนการบริหารจัดการคลังคำศัพท์ โดยมีรายละเอียดดังภาพที่ 3.16



ภาพที่ 3.16 กระบวนการสร้างคำศัพท์ใหม่ (Create Corpus)

จากภาพที่ 3.16 เป็นกระบวนการเพิ่มคำศัพท์ใหม่ โดยประกอบด้วยกระบวนการแยกย่อยทั้งหมด 4 กระบวนการดังนี้ 3.4.1) กระบวนการเพิ่มคำศัพท์ใหม่ 3.4.2) กระบวนการเพิ่มข้อมูลลงในฐานข้อมูลคลังคำศัพท์ 3.4.3) กระบวนการเข้ารหัสข้อมูล และ 3.4.4) กระบวนการเขียนข้อมูลลงใน Text File

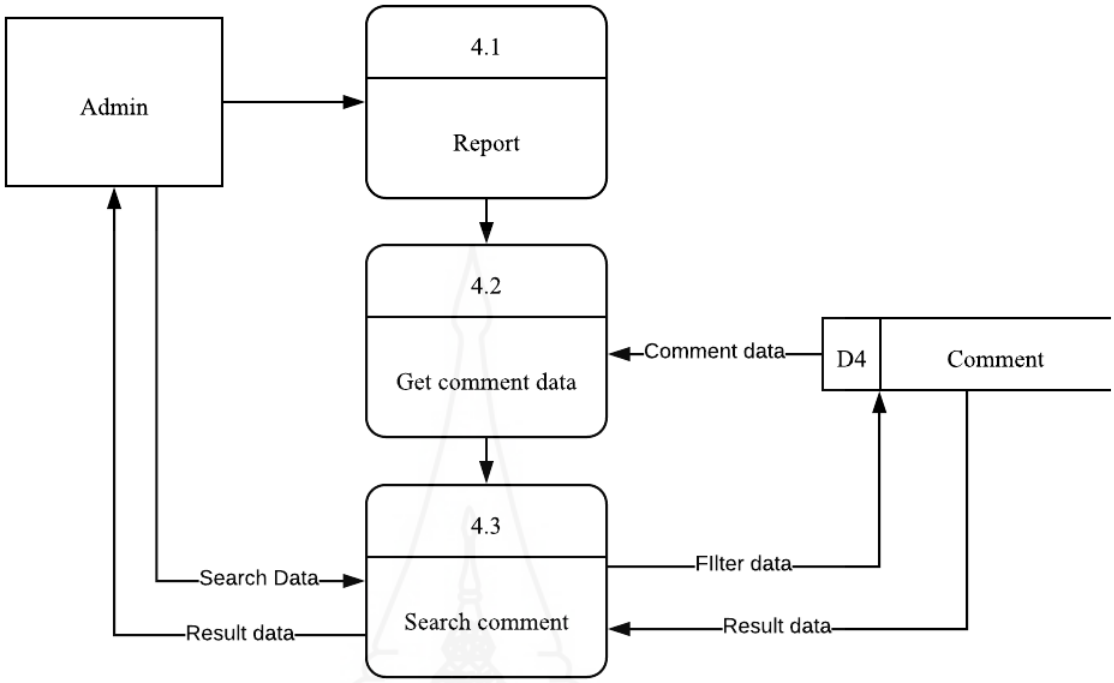
4) กระบวนการลบข้อมูลคลังคำศัพท์ (Delete Corpus) สำหรับกระบวนการนี้เป็นส่วนหนึ่งของกระบวนการบริหารจัดการคลังคำศัพท์ โดยมีรายละเอียดดังภาพที่ 3.17



ภาพที่ 3.17 กระบวนการสร้างคำศัพท์ใหม่ (Create Corpus)

จากภาพที่ 3.17 เป็นกระบวนการลบคำศัพท์ออกจากฐานข้อมูลคลังคำศัพท์ โดยประกอบด้วยกระบวนการแยกย่อยทั้งหมด 4 กระบวนการดังนี้ 3.4.1) กระบวนการเพิ่มคำศัพท์ใหม่ 3.4.2) กระบวนการเพิ่มข้อมูลลงในฐานข้อมูลคลังคำศัพท์ 3.4.3) กระบวนการเข้ารหัสข้อมูล และ 3.4.4) กระบวนการเขียนข้อมูลลงใน Text File

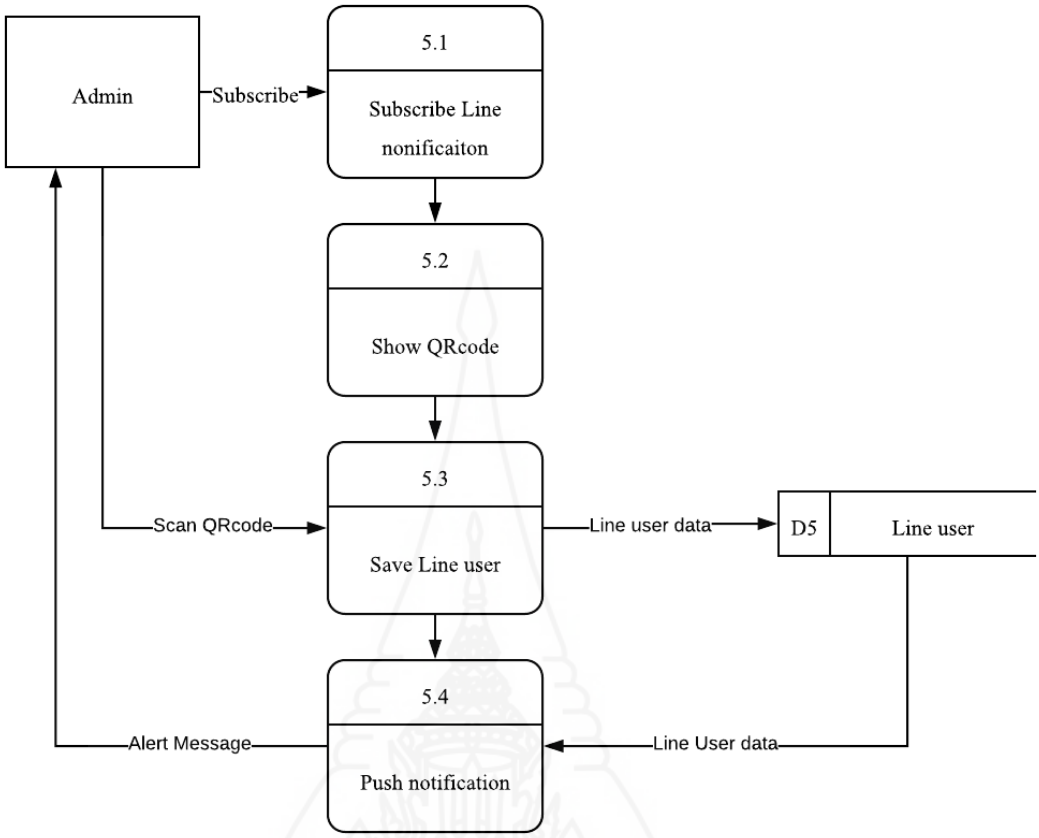
2.4.4 กระบวนการออกรายงานผลการวิเคราะห์ความคิดเห็น (Report) เป็นกระบวนการออกรายงานข้อมูลความคิดเห็นที่เก็บในฐานข้อมูลความคิดเห็น พร้อมผลการวิเคราะห์ความรู้สึก อีกทั้งยังสามารถสืบค้นข้อมูลความคิดเห็นได้อีกด้วย ดังภาพที่ 3.18



ภาพที่ 3.18 กระบวนการออกรายงานผลการวิเคราะห์ความคิดเห็น (Report)

จากภาพที่ 3.18 เป็นกระบวนการออกรายงานผลการวิเคราะห์ความคิดเห็นซึ่งประกอบไปด้วย 3 ขั้นตอนด้วยกันคือ 4.1) กระบวนดูข้อมูลรายงาน 4.2) กระบวนการดึงข้อมูลความคิดเห็นทั้งหมด และ 4.3)กระบวนการค้นหาความคิดเห็นตามเงื่อนไขนั้น ๆ สำหรับแหล่งข้อมูลในกระบวนการทั้งหมดนี้ใช้ข้อมูลจาก ฐานข้อมูลความคิดเห็น และข้อมูลที่ผู้ดูและระบบจะได้รับคือข้อความคิดเห็นและผลลัพธ์การค้นหาข้อมูล

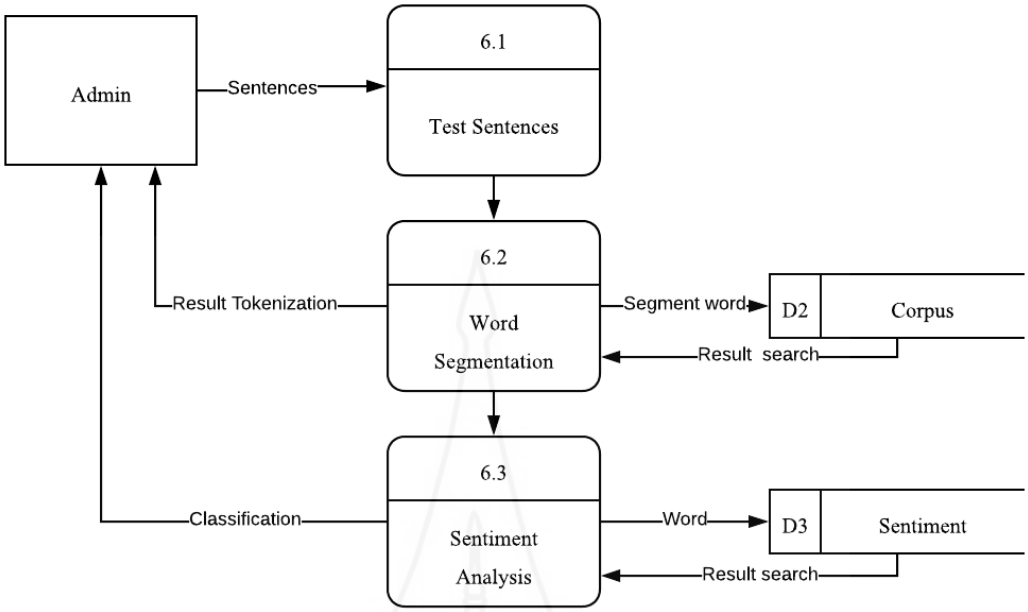
2.4.5 กระบวนการสมัครรับการแจ้งเตือนผ่านทาง Line (Subscribe Line) เป็นกระบวนการสมัครรับการแจ้งเตือนข้อความผ่านทาง Line กรณีที่พบความคิดเห็นเป็นเชิงลบ โดยระบบจะแจ้งข้อความไปยัง Line ของผู้ดูแลระบบที่ได้สมัครรับการแจ้งเตือนเอาไว้ มีขั้นตอนเริ่มต้นจากระบบแสดง QRcode ให้กับผู้ดูแลระบบจากนั้นผู้ดูแลระบบทำการแสกน QRCode ด้วยแอปพลิเคชัน Line ระบบจะได้ ข้อมูลผู้ใช้งาน Line แล้วระบบจะทำการบันทึกข้อมูลเอาไว้เพื่อ ใช้ในการส่งข้อความแจ้งเตือนในลำดับถัดไป สามารถอธิบายรายละเอียดได้ดังภาพ 3.19



ภาพที่ 3.19 กระบวนการสมัครรับการแจ้งเตือนผ่านทาง (Subscribe Line Notification)

จากภาพที่ 3.19 เป็นกระบวนการสมัครรับการแจ้งเตือนผ่านทาง Line ซึ่งประกอบไปด้วย 4 กระบวนการแยกย่อยดังนี้ 5.1) เริ่มต้นกระบวนการสมัครรับการแจ้งเตือนผ่านทาง Line 5.2) กระบวนการนำข้อมูล QRCode มาแสดงผลบนเว็บไซต์ จากนั้นเมื่อผู้ดูระบบทำการแสกน QRCode ข้อมูลจะถูกบันทึกเข้าระบบฐานข้อมูลผู้ใช้งาน Line เพื่อจัดเก็บไว้ใช้ในการส่งข้อความแจ้งเตือนในกระบวนการที่ 5.3) และกระบวนการสุดท้ายคือ 5.4) กระบวนการแจ้งเตือนข้อความไปยังผู้ดูแลระบบ

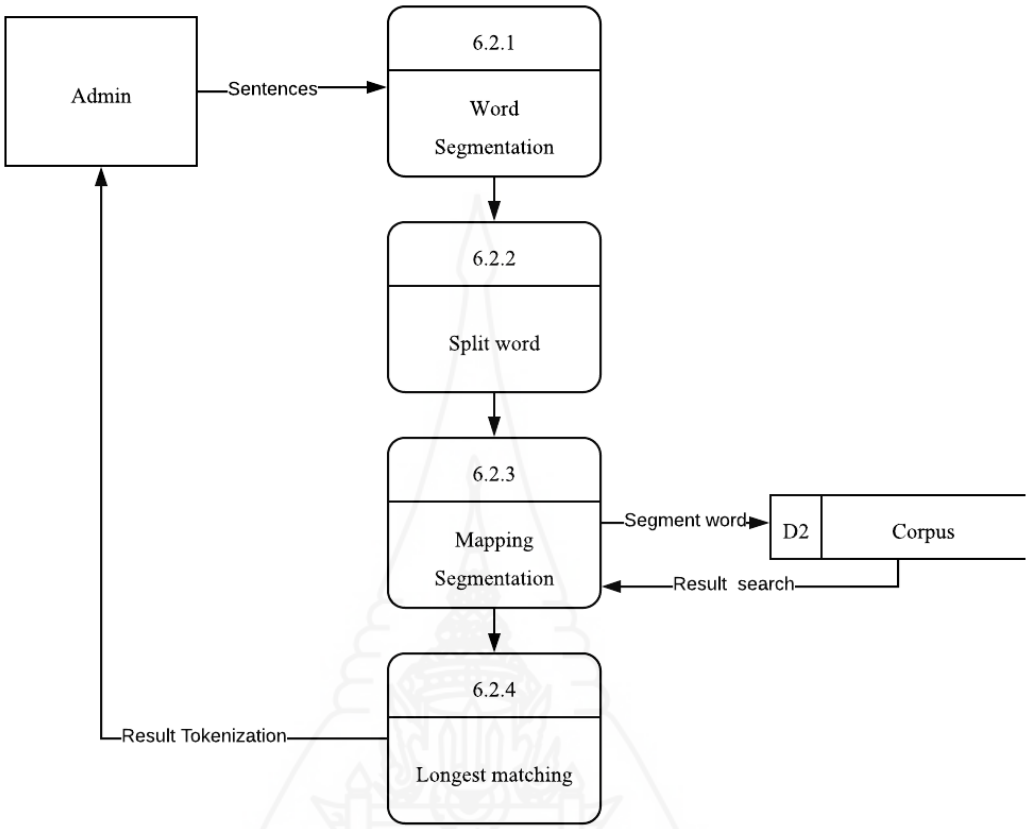
2.4.6 กระบวนการทดสอบข้อความ (Test Sentences) เป็นกระบวนการที่ออกแบบมาเพื่อให้ผู้ดูแลระบบสามารถ นำข้อความที่ต้องการมาทดสอบในระบบโดยมาทดสอบได้ โดยไม่จำเป็นต้องเป็นข้อความที่มีอยู่ในระบบเท่านั้น เป็นข้อความอะไรก็ได้ จะเป็นวิธีการนำเข้าโดยการพิมพ์เข้ามาหรือการคัดลอกมาจากแหล่งข้อมูลอื่น ก็สามารทำได้เช่นเดียวกัน โดยมีรายละเอียดแยกย่อยดังภาพที่ 3.20



ภาพที่ 3.20 กระบวนการทดสอบข้อความ (Test Sentences)

จากภาพที่ 3.20 เป็นกระบวนการทดสอบข้อความ ซึ่งประกอบด้วยขั้นตอนแยกย่อยอยู่ 3 ขั้นตอนด้วยกันคือ 6.1) กระบวนการทดสอบข้อความ 6.2) กระบวนตัดคำ และ 6.3) กระบวนการวิเคราะห์ความรู้สึก ในกระบวนการนี้ ผู้ดูแลระบบสามารถนำข้อความที่ต้องการทดสอบ ด้วยการส่งข้อความนั้นเข้าสู่ระบบจากนั้นระบบจะนำไปตัดคำ โดยใช้หลักการเทียบคำกับฐานข้อมูลคลังคำศัพท์ เมื่อเสร็จสิ้นกระบวนการตัดจะนำคำที่ตัดไปเข้าสู่กระบวนการวิเคราะห์ความรู้สึก โดยการเทียบกับฐานข้อมูลแยกตามประเภทความรู้สึก (Sentiment) ผลลัพธ์ที่ได้จากกระบวนการทั้งหมดนี้คือ ผลลัพธ์การตัดคำ และผลลัพธ์การวิเคราะห์ความรู้สึก

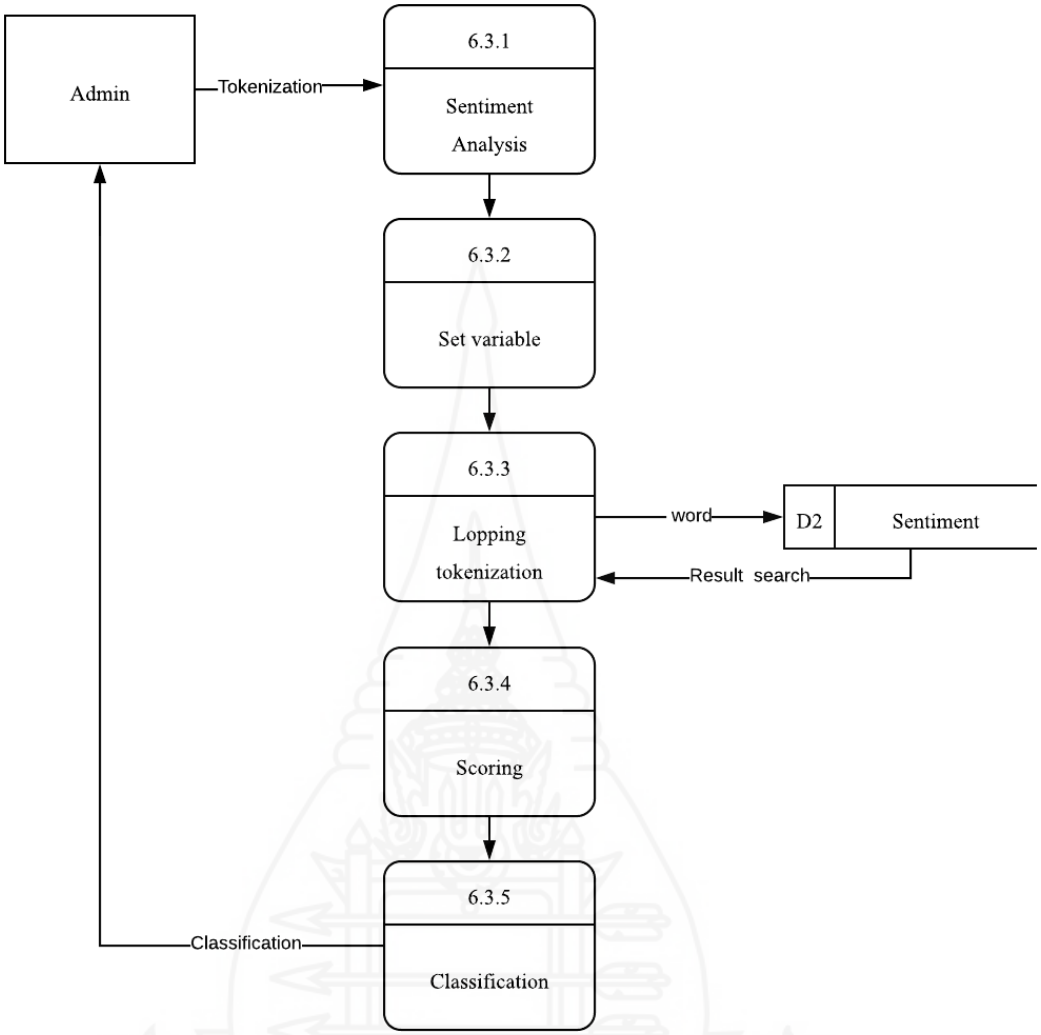
1) กระบวนการตัดคำ (Word Segmentation) กระบวนการนี้เป็นส่วนหนึ่งของกระบวนการทดสอบข้อความในกระบวนการที่ 6.1 ระบบจะดำเนินการโดยเทียบคำกับฐานข้อมูลคลังคำศัพท์ (Corpus) ด้วยวิธีการใช้การจับคู่ที่ยาวที่สุด (Longest Matching) ซึ่งสามารถอธิบายรายละเอียดแยกย่อยดังภาพที่ 3.21



ภาพที่ 3.21 กระบวนการตัดคำ (Word segmentation)

จากภาพที่ 3.21 เป็นการอธิบายขั้นตอนกระบวนการตัดคำ ซึ่งประกอบไปด้วย 4 กระบวนการแยกย่อยดังนี้ 6.2.1) เริ่มต้นกระบวนการตัดคำ 6.2.2) กระบวนการแยกสายอักขระออกจากกัน 6.2.3) ค้นหาสายอักขระโดยเทียบกับการข้อมูลคลังคำศัพท์ และกระบวนการสุดท้าย 6.2.4) เลือกคำที่ยาวที่สุด และส่งออกเป็นผลลัพธ์การตัดคำ

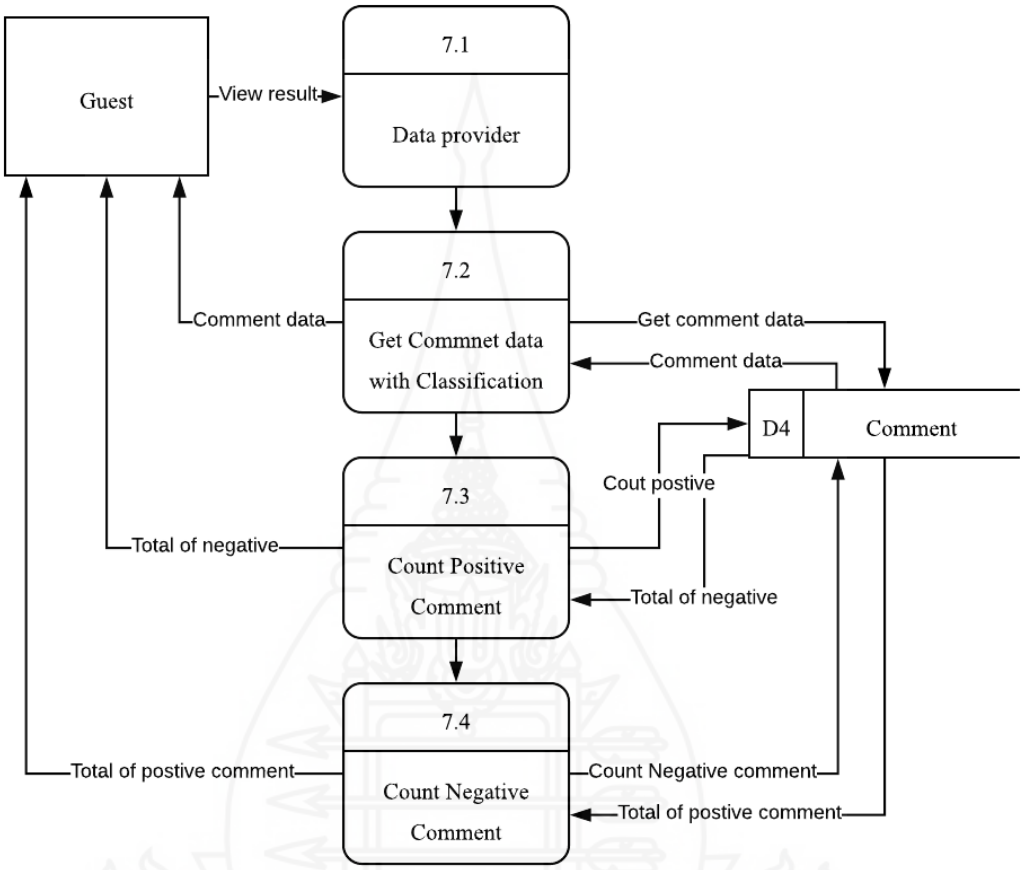
2) กระบวนการวิเคราะห์ความรู้สึก (Sentiment Analysis) กระบวนการนี้เป็นส่วนหนึ่งของกระบวนการทดสอบข้อความในกระบวนการที่ 6.1 กระบวนการวิเคราะห์ความรู้สึกโดยใช้เทคนิคการจำแนกข้อมูลนาอีฟเบย์ (Naive Bayes Classifier) ในกระบวนการนี้จะนำคำที่ถูกตัดไปเทียบกับฐานข้อมูลคลังคำศัพท์ที่แยกตามประเภทความรู้สึก (Sentiment) เพื่อตรวจสอบและคำนวณหาค่าคะแนน หากพบว่าค่าคะแนนในกลุ่มประเภทความรู้สึกใดมีค่ามากที่สุด หมายความว่าข้อความที่นำเข้ามาจะเป็นประเภทนั้น หลังจากนั้นจะทำการบันทึกผลที่ได้จากการวิเคราะห์กลับไปที่ฐานข้อมูลความคิดเห็น ซึ่งสามารถอธิบายรายละเอียดแยกย่อยดังภาพที่ 3.22



ภาพที่ 3.22 กระบวนการวิเคราะห์ความรู้สึก (Sentiment Analysis)

จากภาพที่ 3.22 เป็นการอธิบายขั้นตอนกระบวนการวิเคราะห์ความรู้สึก ซึ่งประกอบไปด้วย 5 กระบวนการแยกย่อยดังนี้ 6.3.1) เริ่มต้นกระบวนการวิเคราะห์ความรู้สึก 6.3.2) ตั้งค่าตัวแปร 6.3.3) วนลูปคำที่ถูกตัดโดยเทียบแต่ละคำกับฐานข้อมูลแยกตามประเภทความรู้สึก (Sentiment) 6.3.4) หาค่าคะแนน และกระบวนการสุดท้าย 6.3.5) เรียงลำดับค่าคะแนนพร้อมส่งออกผลลัพธ์การวิเคราะห์ความรู้สึก

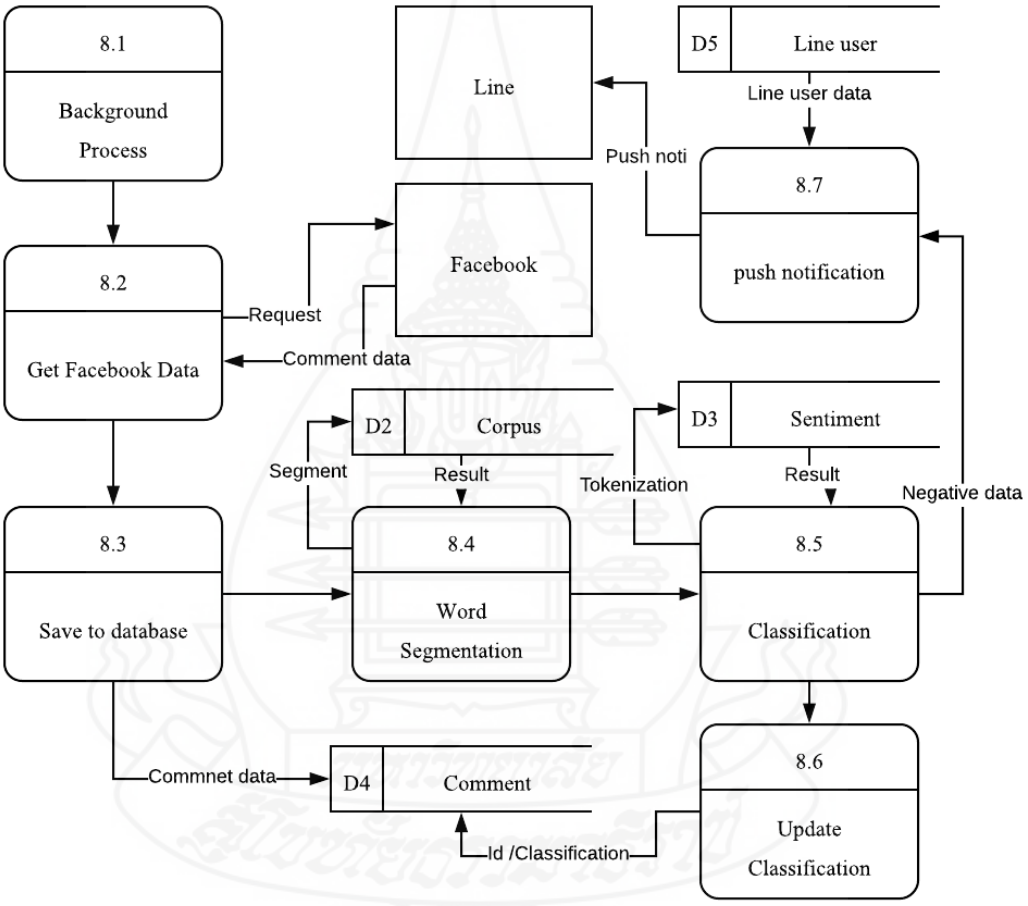
2.4.7 กระบวนการให้บริการข้อมูล (Data Provider) เป็นกระบวนการที่ออกแบบมาเพื่อส่งข้อมูลผลการวิเคราะห์ความรู้สึกนำไปแสดงผลในส่วนติดต่อกับผู้ใช้งานทั่วไป (Guest) ซึ่งสามารถอธิบายรายละเอียดแยกย่อยดังภาพที่ 3.23



ภาพที่ 3.23 กระบวนการให้บริการข้อมูล (Data Provider)

จากภาพที่ 3.23 เป็นการอธิบายขั้นตอนกระบวนการให้บริการข้อมูลสำหรับส่วนติดต่อผู้ใช้งานทั่วไป (Front-end) ซึ่งประกอบไปด้วย 4 กระบวนการแยกย่อยดังนี้ 7.1) เริ่มต้นกระบวนการ 7.2) กระบวนการรวมรวมข้อมูลความคิดเห็นพร้อมผลการวิเคราะห์ความรู้สึกจากฐานข้อมูลความคิดเห็น 7.3) กระบวนการนับความคิดเห็นที่เป็นความรู้สึกเชิงบวก และ 7.4) กระบวนการนับความคิดเห็นที่เป็นความรู้สึกเชิงลบ ข้อมูลทั้งหมดที่ส่งออกไปยังผู้ใช้งานทั่วไป (Guest) ประกอบด้วย จำนวนความคิดเห็นเชิงบวก จำนวนคิดเห็นเชิงลบและ ความคิดเห็นพร้อมทั้งผลลัพธ์การวิเคราะห์ความรู้สึก

2.4.8 กระบวนการทำงานเบื้องหลัง (Background Process) เป็นกระบวนการที่คอยทำงานอยู่เบื้องหลัง ที่จะทำหน้าที่รวบรวมข้อมูลจากเฟซบุ๊ก บันทึกข้อมูลที่ได้มา นำเข้าสู่กระบวนการตัดคำ แล้วนำคำที่ผ่านกระบวนการตัดคำเข้าสู่กระบวนการวิเคราะห์ความรู้สึก เสร็จแล้วบันทึกผลลัพธ์การวิเคราะห์ความรู้สึก หากพบว่ามีความคิดเห็นที่ระบบวิเคราะห์ออกมาว่าเป็นความคิดเห็นเชิงลบแล้ว ระบบจะส่งข้อมูลไปยังกระบวนการแจ้งเตือน เพื่อทำการแจ้งผู้ดูแลระบบที่เคยสมัครรับการแจ้งเตือนผ่านทาง Line ซึ่งสามารถอธิบายรายละเอียดดังภาพที่ 3.24

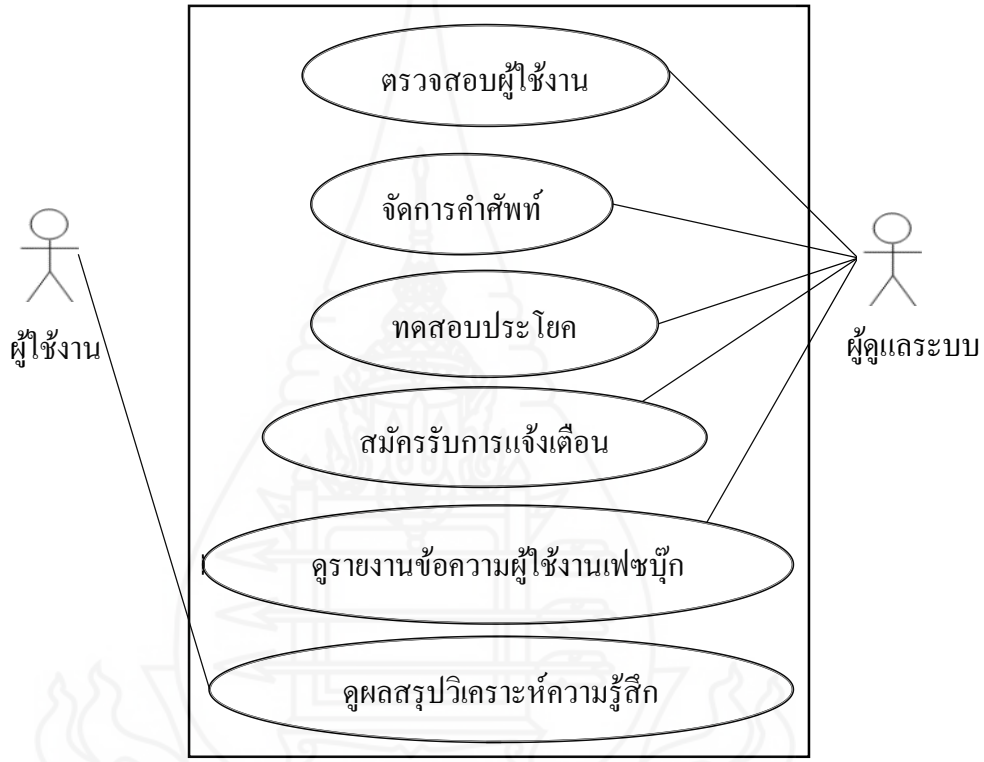


ภาพที่ 3.24 กระบวนการทำงานเบื้องหลัง (Background Process)

จากภาพที่ 3.24 เป็นการอธิบายขั้นตอนกระบวนการทำงานเบื้องหลัง ซึ่งประกอบไปด้วย 7 กระบวนการแยกย่อยดังนี้ 8.1) เริ่มต้นกระบวนการทำงาน 8.2) กระบวนการสืบค้นข้อมูลจากเฟซบุ๊ก 8.3) กระบวนการบันทึกข้อมูลที่ได้จากเฟซบุ๊กมาบันทึกลงในฐานข้อมูลความคิดเห็น 8.4) กระบวนการตัดคำโดยการเทียบคำกับฐานข้อมูลคลังคำศัพท์ 8.5) กระบวนการวิเคราะห์ความรู้สึก

8.6) กระบวนการบันทึกผลการวิเคราะห์ความรู้สึก และ 8.7) กระบวนการแจ้งเตือนกรณีที่พบความรู้สึกเป็นเชิงลบ

2.5 การออกแบบแผนภาพยูสเคสไดอะแกรม (Use Case Diagram) เพื่อแสดงขอบเขตสิทธิ์การใช้งานระบบ ซึ่งประกอบด้วยผู้ใช้ระบบ (Actor) ดังนี้ คือ ผู้ดูแลระบบ (Admin) และผู้ใช้งาน (User) ดังแสดงในภาพที่ 3.25



ภาพที่ 3.25 ยูสเคสไดอะแกรมสำหรับการพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์

2.5.1 คำอธิบายยูสเคส (Use Case Description) จากยูสเคสไดอะแกรมที่ได้จากภาพที่ 3.25 สามารถนำมากำหนดรายละเอียดย่อยในแต่ละฟังก์ชันได้ดังนี้

1) ยูสเคส ตรวจสอบผู้ใช้งาน เป็นการตรวจสอบสิทธิ์ผู้เข้าใช้งานของระบบโดยผู้ดูแลระบบต้องทำการล็อกอิน (Login) โดยป้อน Username และ Password ที่ถูกต้อง ดังคำอธิบายในตารางที่ 3.3

ตารางที่ 3.3 คำอธิบายยูสเคส ตรวจสอบผู้ใช้งาน

Use Case Name:	**ตรวจสอบผู้ใช้งาน**
Actor:	ผู้ดูแลระบบ
Description:	แสดงสิทธิ์การเข้าใช้งานระบบ
Normal Course:	ผู้ดูแลระบบทำการล็อกอิน โดยป้อน Username และ Password ที่ถูกต้องจึงสามารถเข้าใช้งานระบบได้
Alternate Course:	ผู้ดูแลระบบที่ป้อน Username และ Password ไม่ถูกต้องจะไม่สามารถเข้าใช้งานระบบได้
Precondition:	-
Post-condition:	-
Assumptions:	-

2) ยูสเคส จัดการคำศัพท์ เป็นการจัดการคำศัพท์ในพจนานุกรมคลังคำศัพท์ (Corpus) โดยสามารถเพิ่มพจนานุกรม และแก้ไขคำศัพท์ในพจนานุกรม ดังคำอธิบายในตารางที่ 3.4

ตารางที่ 3.4 คำอธิบายยูสเคส จัดการคำศัพท์

Use Case Name:	**จัดการคำศัพท์**
Actor:	ผู้ดูแลระบบ
Description:	ผู้ดูแลระบบสามารถนำเข้าพจนานุกรมคำศัพท์มาเก็บไว้ในฐานข้อมูลของระบบ พร้อมทั้งเพิ่มและแก้ไขคำศัพท์ในพจนานุกรม
Normal Course:	ผู้ดูแลระบบจะต้องมีการแสดงสิทธิ์การเข้าใช้งานระบบเท่านั้นจึงจะสามารถจัดการคำศัพท์ได้

ตารางที่ 3.4 (ต่อ)

Alternate Course:	ผู้ดูแลระบบที่ไม่แสดงสิทธิ์การเข้าใช้งานไม่สามารถใช้งานระบบได้
Precondition:	-
Post-condition:	-
Assumptions:	-

3) ยูสเคส ทดสอบประโยค เป็นการนำประโยคเข้าทดสอบการตัดคำและการแสดงสถานะของคำในประโยค เพื่อทดสอบความถูกต้อง พร้อมทั้งสามารถแก้ไขและเพิ่มสถานะของคำที่ตัดจากประโยคได้ ดังคำอธิบายในตารางที่ 3.5

ตารางที่ 3.5 คำอธิบายยูสเคส ทดสอบประโยค

Use Case Name:	**ทดสอบประโยค**
Actor:	ผู้ดูแลระบบ
Description:	สามารถนำประโยคเข้าทดสอบการตัดคำและการแสดงสถานะของคำในประโยค เพื่อทดสอบความ ถูกต้อง พร้อมทั้งสามารถแก้ไขและเพิ่มสถานะของคำที่ตัดจากประโยคได้
Normal Course:	ผู้ดูแลระบบจะต้องมีสิทธิ์การเข้าใช้งานระบบเท่านั้นจึงจะสามารถทดสอบประโยคได้
Alternate Course:	ผู้ดูแลระบบที่ไม่แสดงสิทธิ์การเข้าใช้งานไม่สามารถใช้งานระบบได้
Precondition:	-
Post-condition:	-
Assumptions:	-

4) ยูสเคส รายงานข้อความผู้ใช้งานเฟซบุ๊ก แสดงกระแสข้อมูลการแสดงความคิดเห็นของผู้ใช้งานที่เข้าสู่ระบบแบบเรียลไทม์ (Real-time) โดยผ่านการวิเคราะห์ข้อมูลแล้ว ดังคำอธิบายในตารางที่ 3.6

ตารางที่ 3.6 คำอธิบายยูสเคส ดูรายงานข้อความผู้ใช้งานเฟซบุ๊ก

Use Case Name:	**ดูรายงานข้อความผู้ใช้งานเฟซบุ๊ก**
Actor:	ผู้ดูแลระบบ
Description:	แสดงกระแสข้อมูลการแสดงความคิดเห็นของผู้ใช้งานที่เข้าสู่ระบบแบบเรียลไทม์ (Real-time) ที่ผ่านการวิเคราะห์ข้อมูลแล้ว
Normal Course:	ผู้ดูแลระบบจะต้องมีการแสดงสิทธิ์การเข้าใช้งานระบบเท่านั้นจึงจะสามารถดูรายงานข้อความผู้ใช้งานเฟซบุ๊กได้
Alternate Course:	ผู้ดูแลระบบที่ไม่แสดงสิทธิ์การเข้าใช้งานไม่สามารถใช้งานระบบได้
Precondition:	-
Post-condition:	-
Assumptions:	-

5) ยูสเคส ข้อความแจ้งเตือน แสดงข้อความผ่านไลน์ (Line) เมื่อมีการแสดงความคิดเห็นที่เป็นความรู้สึกเชิงลบ ดังคำอธิบายในตารางที่ 3.7

ตารางที่ 3.7 คำอธิบายยูสเคส สมัครรับการแจ้งเตือน

Use Case Name:	**สมัครรับการแจ้งเตือน**
Actor:	ผู้ดูแลระบบ
Description:	แสดงข้อความผ่านไลน์เมื่อมีการแสดงความคิดเห็นที่เป็นความรู้สึกเชิงลบ

ตารางที่ 3.7 (ต่อ)

Normal Course:	ผู้ดูแลระบบจะต้องมีการแสดงสิทธิ์การเข้าใช้งานระบบเท่านั้นจึงจะสามารถรับข้อความแจ้งเตือนได้
Alternate Course:	ผู้ดูแลระบบที่ไม่แสดงสิทธิ์การเข้าใช้งานไม่สามารถใช้งานระบบได้
Precondition:	-
Post-condition:	-
Assumptions:	-

6) ยูสเคส ผลสรุปวิเคราะห์ความรู้สึก แสดงกระแสข้อมูลการแสดงความคิดเห็นของผู้ใช้งานที่เข้าสู่ระบบแบบเรียลไทม์ (Real-time) โดยผ่านการวิเคราะห์ข้อมูลแล้ว ดังคำอธิบายในตารางที่ 3.8

ตารางที่ 3.8 คำอธิบายยูสเคส ดูผลสรุปวิเคราะห์ความรู้สึก

Use Case Name:	**ผลสรุปวิเคราะห์ความรู้สึก**
Actor:	ผู้ดูแลระบบ และ ผู้ใช้งาน
Description:	แสดงผลข้อความที่วิเคราะห์ความรู้สึก (Sentiment Analysis) พร้อมทั้งสถิติความรู้สึกเชิงบวก (Positive) และความรู้สึกเชิงลบ (Negative)
Normal Course:	-
Alternate Course:	-
Precondition:	-
Post-condition:	-
Assumptions:	-

3. ข้อมูลและกลุ่มตัวอย่างที่ใช้ในงานวิจัย

ระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ (Real-time) ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูล นาอีฟ เบย์ (Naive Bayes Classifier) สำหรับภาษาไทย ใช้ข้อมูลจากเพจเฟซบุ๊กมหาวิทยาลัยสุโขทัยธรรมาธิราช STOUniversity (https://fb.com/STOUniversity/) ซึ่งเป็นข้อมูลการแสดงความคิดเห็นของนักศึกษาที่ติดตาม เฟซบุ๊กเพจของมหาวิทยาลัยสุโขทัยธรรมธิราช จำนวน 241,935 คนได้เริ่มเปิดเฟซบุ๊กเพจเมื่อวันที่ 7 สิงหาคม 2553 เป็นต้นมา โดยมีการเก็บข้อมูลการแสดงความคิดเห็นของนักศึกษาที่ติดตามเฟซบุ๊กเพจในรูปแบบฐานข้อมูล และข้อมูลอื่น ๆ ที่ใช้งานในระบบ ซึ่งมีรายละเอียด ดังตารางที่ 3.9

ตารางที่ 3.9 รายละเอียดการจัดเก็บข้อมูล

ลำดับที่	ชื่อคอลัมน์	ประเภท	รายละเอียด
1	id	Integer	ลำดับรายการข้อมูลแสดงความคิดเห็น
2	comment_id	Integer	รหัสข้อความความคิดเห็นจากเฟซบุ๊ก
3	comment	String	ข้อความแสดงความคิดเห็น
4	created	Datetime	วันและเวลาการแสดงความคิดเห็น
5	created_at	Datetime	วันและเวลาในการบันทึกข้อมูล
6	updated_at	Datetime	วันและเวลาในการปรับปรุงข้อมูล

เพื่อนำข้อมูลที่ได้มาใช้ในการแสดงผลข้อมูลการแสดงความคิดเห็นภายในระบบสำหรับการวิเคราะห์ความรู้สึกของนักศึกษา

ตารางที่ 3.10 ตัวอย่างข้อมูลการแสดงความคิดเห็นที่จัดเก็บไว้ในฐานข้อมูลระบบ

id	comment_id	comment	created
1	1687634667923627	เป็นเพื่อนรักของป้าแมวคะ สามีเป็น...	2017-12-15 09:32:36
2	1687525604601200	ขอบพระคุณมากค่ะ	2017-12-15 07:30:31
3	1687550294598731	นามสกุลคุ้นๆ จัง ^^	2017-12-15 07:50:22

ตารางที่ 3.10 (ต่อ)

id	comment_id	comment	created
4	1687604971259930	แชร์ได้ไหมครับท่านอาจารย์	2017-12-15 08:49:59
5	1687769117910182	นับถือ respect มากค้าบ	2017-12-15 12:24:39
6	238585183345934	สนใจค่ะ	2017-11-29 16:37:07
7	1629733167085285	เราจะก้าวไปด้วยกันนะครับ	2017-11-29 13:03:18
8	1557064751043275	ป.โท เปิดหรือยังค่ะ	2017-11-29 11:44:35

ตารางที่ 3.11 รายละเอียดการจัดเก็บข้อมูลคลังคำศัพท์

ลำดับที่	ชื่อคอลัมน์	ประเภท	รายละเอียด
1	id	Integer	ลำดับรายการข้อมูลคำศัพท์
2	word	String	ข้อมูลคำศัพท์
3	created_at	Datetime	วันและเวลาการสร้างข้อมูล
4	updated_at	Datetime	วันและเวลาการปรับปรุงข้อมูล

ตารางที่ 3.12 ตัวอย่างข้อมูลคลังคำศัพท์ที่จัดเก็บไว้ในฐานข้อมูลระบบ

id	word	created_at	updated_at
1	กกกร	2017-10-24 16:15:46	2017-10-24 16:15:46
2	กกช.	2017-10-24 16:15:46	2017-10-24 16:15:46
3	กกจ.	2017-10-24 16:15:46	2017-10-24 16:15:46
4	กกช.	2017-10-24 16:15:46	2017-10-24 16:15:46
5	กกต.	2017-10-24 16:15:46	2017-10-24 16:15:46
6	กงการ	2017-10-24 16:15:46	2017-10-24 16:15:46
7	กงจักร	2017-10-24 16:15:46	2017-10-24 16:15:46
8	กงฉาก	2017-10-24 16:15:46	2017-10-24 16:15:46
9	กงทอง	2017-10-24 16:15:46	2017-10-24 16:15:46

ตารางที่ 3.13 รายละเอียดการจัดเก็บฐานข้อมูลคลังคำศัพท์แยกตามประเภทความรู้สึก

ลำดับที่	ชื่อคอลัมน์	ประเภท	รายละเอียด
1	id	Integer	ลำดับรายการข้อมูลคำศัพท์
2	word	String	ข้อมูลคำศัพท์
3	word_type	String	ประเภทความรู้สึก
4	created_at	Datetime	วันและเวลาการสร้างข้อมูล
5	updated_at	Datetime	วันและเวลาการปรับปรุงข้อมูล

ตารางที่ 3.14 ตัวอย่างข้อมูลคลังคำศัพท์ ประเภทความรู้สึกเชิงบวกที่จัดเก็บไว้ในฐานข้อมูลระบบ

id	word	word_type	created_at	updated_at
1	ดี	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
2	ขอบคุณ	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
3	สู้เสมอ	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
4	โคตรสนุก	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
5	สุดยอด	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
6	น่ารัก	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
7	ขอบคุณ	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
8	รอยยิ้ม	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
9	รักมาก	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
10	คิดถึงมาก	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
11	ดีใจ	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
12	มิตรภาพดี	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
13	ที่รัก	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
14	ชอบจัง	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
15	สวยงาม	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
16	ดีใจด้วย	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30
17	ทรงพระเจริญ	Positive	2017-10-24 16:01:30	2017-10-24 16:01:30

ตารางที่ 3.15 ตัวอย่างข้อมูลคลังคำศัพท์ ประเภทความรู้สึกเชิงลบที่จัดเก็บไว้ในฐานข้อมูลระบบ

id	word	word_type	created_at	updated_at
1	ไม่	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
2	แย่ว่ะ	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
3	ปวดหัว	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
4	เลว	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
5	ย่ำแย่	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
6	อึดอัด	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
7	ขี้เกียจ	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
8	ไม่ชอบ	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
9	เสียความรู้สึก	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
10	ทุเรศ	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
11	เดรัจฉาน	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
12	ตำหนิ	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
13	เสียน้ำตา	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
14	หัวร้อน	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
15	ยาก	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
16	อ่อนเพลีย	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
17	ฟุ้งซ่าน	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
18	ไม่สร้างสรรค์	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
19	เลิกเถอะ	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
20	แพง	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
21	เสื่อม	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
22	เครียด	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
23	เหี้ย	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
24	เบื่อ	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26
25	เหยียบซ้ำ	Negative	2017-10-24 16:01:26	2017-10-24 16:01:26

ตารางที่ 3.16 ตัวอย่างข้อมูลคลังคำศัพท์ ประเภทความรู้สึกเป็นกลาง

id	word	word_type	created_at	updated_at
1	เฉย	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
2	นิ่ง	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
3	เสมอ	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
4	ตกลง	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
5	อ่อ	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
6	สัมภาษณ์	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
7	มาตรการ	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
8	รับรู้	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
9	จับ	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
10	เกาะ	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
11	พอกัน	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
12	ธรรมชาติ	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
13	ดำรง	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
14	เผื่อ	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
15	เปิดเผย	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
16	บอก	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
17	พิสูจน์	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
18	กลาง	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
19	เป็นกลาง	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
20	เว้น	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
21	ปล่อย	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
22	สถานะ	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
23	สภาวะ	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
24	ธรรมดา	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26
25	เท่าเทียม	Neutral	2017-10-24 16:01:26	2017-10-24 16:01:26

ตารางที่ 3.17 รายละเอียดการจัดเก็บฐานข้อมูลความคิดเห็นที่ผ่านกระบวนการตัดคำภาษาไทย

ลำดับที่	ชื่อคอลัมน์	ประเภท	รายละเอียด
1	id	Integer	ลำดับรายการข้อมูลคำศัพท์
3	comment_split	String	ข้อมูลความคิดเห็นที่ตัดแล้ว
4	word_type	String	ข้อมูลผลการวิเคราะห์ความรู้สึก
5	created_at	Datetime	วันและเวลาการสร้างข้อมูล

ตารางที่ 3.18 ตัวอย่างข้อมูลการตัดคำและผลลัพธ์การวิเคราะห์ความรู้สึก

id	comment_split	word_type	created_at
1	ยินดี|ที่|ได้|รู้|จัก	Positive	2017-12-15 17:44:02
2	อ.|น้อง|สวย|มากๆ|ค่ะ	Positive	2017-12-15 17:44:02
3	น่าจะ|มี|ออนไลน์|นะ	Neutral	2017-12-15 17:44:02
4	ภูมิใจ|ที่|ได้|เรียน|ที่นี่|คะ	Positive	2017-12-15 17:44:07
5	เรา|จะ|ก้าว|ไป|ด้วยกัน|นะ|ครับ	Positive	2017-12-15 17:44:06
6	สู้ๆ|ครับผม|ศิษฐ์|เก่า|เอาใจช่วย	Positive	2017-12-15 17:44:06
7	สู้ๆ|นะ|ค่ะ|ไป|ให้|ถึง|ฝัน	Positive	2017-12-15 17:44:07
8	กด|แล้ว|มันerror	Negative	2017-12-15 17:44:08
9	ทำไม|ส่ง|ทีไร|ก็|error	Negative	2017-12-15 17:44:08
10	ตรง|กับ|วันเกิด|พอดี	Positive	2017-12-15 17:44:11
11	สาธุ|ครับ	Positive	2017-12-15 17:44:11
12	รายละเอียด|ค่าใช้จ่าย|ค่า|หน่วย|กิจ|ฯลฯ|อยู่|ที่|ใบสมัคร|พร้อม|คู่มือ|711|ใกล้|บ้าน|ท่าน	Neutral	2017-12-15 17:44:11
13	กำลัง|เรียน|ป.ตรี|บัญชี|อยู่|ปี|2|ถ้า|จะ|เรียน|เรียน|ควบคู่|ไป|ด้วยกัน|ได้มั้ย|ค่ะ	Neutral	2017-12-15 17:44:12
14	ผลิต|หนังสือ|ชุดวิชา|ที่|ให้ลงทะเบียนเรียน|ให้เสร็จก่อน|ดีมั๊ย|ก่อน|จะ|เปิดรับ|นักศึกษา|	Negative	2017-12-15 17:44:13

ตารางที่ 3.19 รายละเอียดการจัดเก็บฐานข้อมูลผู้ดูแลระบบ

ลำดับที่	ชื่อคอลัมน์	ประเภท	รายละเอียด
1	id	Integer	ลำดับรายการข้อมูล
3	name	String	ชื่อบัญชีผู้ดูแลระบบ
4	email	String	อีเมล
5	password	String	รหัสผ่านป็นข้อมูลเข้ารหัส
6	created_at	Datetime	วันและเวลาการสร้างข้อมูล

ตารางที่ 3.20 ตัวอย่างข้อมูลผู้ดูแลระบบที่จัดเก็บไว้ในฐานข้อมูลระบบ

id	name	email	password	created_at
1	admin	admin@admin.com	$2y$10$f/icLpXTgZJTRridM4eyieLM0lyqEwcttT0yksj4ANgYuDI/IZ8/a	2017-12-15 02:35:40
2	samark	samarkchsngn@gmail.com	$2y$10$51dvnj9nuX/9cPaDFVx2EukOiXfYW2Ul4J6XjM9wJP34OwG9VF5ZW	2018-04-20 04:51:33

4. การพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์

การพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ (Real-time) ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ (Naive Bayes Classifier) สำหรับภาษาไทย จากขั้นตอนการออกแบบและวิเคราะห์ระบบ ทำให้ทราบขั้นตอนของกระบวนการทำงานของระบบงาน โดยผู้วิจัยได้พัฒนาระบบในรูปแบบของเว็บไซต์ (Web Site) พัฒนาด้วยภาษา PHP บน Laravel Framework 5.3 ใช้ MySQL เป็นระบบจัดการฐานข้อมูล และใช้ NodeJS บนภาษา JavaScript สำหรับการทำงานของข้อมูลแบบเรียลไทม์ (Real-time) ซึ่งแบ่งออกเป็นสองส่วนหลักด้วยกันคือ

4.1 ระบบบริหารจัดการข้อมูล (Backend) คือระบบบริหารจัดการคลังข้อมูลคำศัพท์ (Corpus) รายงานผลสืบค้นความคิดเห็นของผู้ติดตามเพจ เพื่อให้ผู้ดูแลระบบสามารถบริหารจัดการ

ข้อความคิดเห็นที่ผู้ติดตามเฟซบุ๊กเพจพร้อมทั้งยังมีเครื่องมือที่ใช้ในการจัดการคลังคำศัพท์ ระบบถูกออกแบบให้ผู้ดูแลระบบสามารถเพิ่มคำศัพท์ใหม่ ๆ ด้วยความง่าย ไม่จำเป็นต้องมีทักษะในการเขียนโปรแกรมก็สามารถทำได้ ในส่วนนี้เองทำให้ระบบมีความยืดหยุ่นสูงและยังช่วยเพิ่มประสิทธิภาพในการตัดคำและวิเคราะห์ความรู้สึก รวมถึงการเรียนรู้คำศัพท์ใหม่ ๆ ของระบบให้สอดคล้องกันสถานการณ์ปัจจุบัน และยังช่วยให้ระบบมีความแม่นยำในการตัดคำมากขึ้นอีกด้วย

4.1.1) การพัฒนาระบบตัดคำภาษาไทย ในการตัดคำภาษาไทยผู้วิจัยเลือกใช้ THSplitLib ในการตัดคำภาษาไทย ซึ่งทำงานบนภาษา PHPโดยการเทียบคำกับฐานข้อมูลคลังคำศัพท์ (Corpus) ด้วยวิธีการใช้การจับคู่ที่ยาวที่สุด (Longest Matching) ในส่วนของระบบงานส่วนผู้นี้ผู้วิจัยได้พัฒนาเพิ่มเติมจากเดิมคือ เพิ่มการบันทึกข้อมูลลงไปในฐานข้อมูลที่เป็น Mysql ซึ่งเดิมทีใช้รูปแบบการบันทึกข้อมูลอยู่ใน Text File ในรูปแบบการเข้ารหัสไฟล์เอาไว้ ซึ่งการแก้ไขทำได้ยาก มีโอกาสที่จะเกิดความผิดพลาดสูง หากมีการไขไฟล์ดังกล่าวโดยที่ไม่มีความรู้ด้านการเขียนโปรแกรม และจะส่งกระทบต่อระบบทำให้ระบบไม่สามารถทำงานได้เป็นปรกติ โดยวัตถุประสงค์หลักของการเพิ่มคุณลักษณะ (Feature) นี้เข้าไปก็เพื่อให้ บุคคลที่ไม่จำเป็นต้องมีความรู้ด้านการเขียนโปรแกรมก็สามารถดำเนินการแก้ไขข้อมูล ฐานข้อมูลคลังคำศัพท์ได้ด้วยตัวเอง ด้วยวิธีการที่ง่ายผ่านเครื่องมือระบบบริหารจัดการข้อมูล (Backend) ที่ผู้วิจัยได้ออกแบบไว้ อีกทั้งยังช่วยให้เพิ่มข้อมูลคำศัพท์ที่เกิดใหม่ได้ทันทีที่ต้องการ

4.1.2) การพัฒนาระบบวิเคราะห์ความรู้สึก ในส่วนนี้ผู้วิจัยเลือกใช้ Open Source ชื่อว่า PhpInsight เป็นเครื่องมือจำแนกข้อความออกเป็นประเภทต่าง ๆ ตามที่ต้องการ ด้วยเทคนิคการจำแนกข้อมูลนาอีฟเบย์ (Naive Bayes Classification) พัฒนาขึ้นมาบนภาษา PHP ซึ่งเหมาะสมกับงานวิจัยชิ้นนี้ แต่มีปัญหาอย่างหนึ่งคือ Open Source ตัวนี้ไม่รองรับการทำงานภาษาไทย เดิมทีความสามารถของ Open Source ตัวนี้คือการนำข้อความภาษาอังกฤษเข้าไปวิเคราะห์ความรู้สึก และผลลัพธ์ที่ได้คือผลการวิเคราะห์ข้อมูลเป็นเชิงบวก เชิงลบ หรือเป็นกลางแต่รองรับการตัดคำได้เฉพาะภาษาอังกฤษเท่านั้น ดังนั้นผู้วิจัยจึงได้พัฒนาเพิ่มเติมคือ แก้ไขส่วนการตัดคำโดยใช้เทคนิคในข้อ 4.12) คือให้ THSplitLib เป็นตัวตัดคำภาษาไทย จากนั้นก็ส่งเข้ากระบวนการวิเคราะห์ความรู้สึกด้วย PhpInsight เพื่อจำแนกข้อความออกมา หลังจากเสร็จแล้วทำการบันทึกผลการวิเคราะห์ความรู้สึกลงในฐานข้อมูลความคิดเห็นตามลำดับ

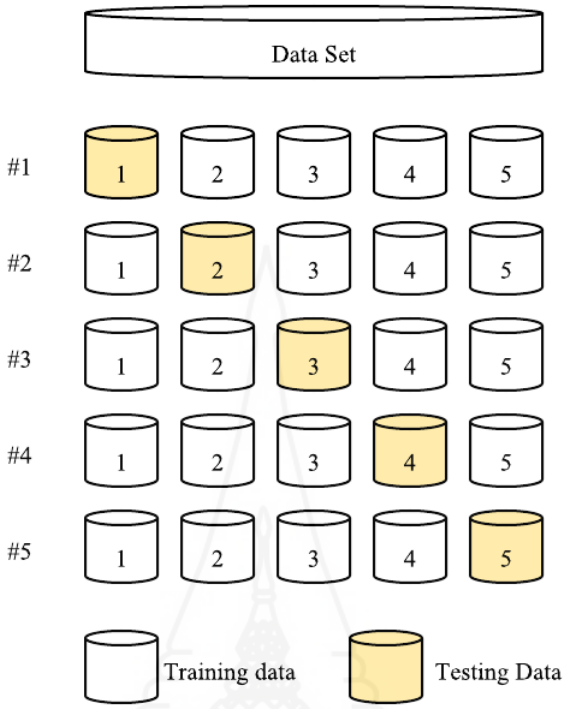
อีกส่วนหนึ่งที่ผู้วิจัยพัฒนาเพิ่มเติมขึ้นมาคือ ส่วนของการเขียนและอ่านไฟล์ ที่ใช้ในการจำแนกข้อมูล เช่นเดียวกับการตัดคำคือ ข้อมูลที่ใช้ในการจำแนกประเภทความรู้สึก ถูกเก็บไว้ในรูปแบบของ Text File ที่เข้ารหัสเอาไว้ ซึ่งการแก้ไขข้อมูลนั้นทำได้ยาก หากไม่มีความรู้ด้านการเขียนโปรแกรม ดังนั้นผู้วิจัยจึงได้พัฒนา นำข้อมูลส่วนนี้ไปจัดเก็บไว้ในฐานข้อมูล Mysql ซึ่งจะทำ

ให้ผู้ดูแลระบบสามารถ บริหารจัดการคำคลังศัพท์ได้ง่าย ผ่านทางเครื่องมือที่ผู้วิจัยได้ออกแบบเอาไว้ แต่ในกระบวนการอ่านไฟล์ก่อนนำไปใช้ ผู้วิจัยยังคงไว้ซึ่งกระบวนการเดิมคือให้ อ่านข้อมูลจาก Text File เช่นเดิม แต่ Text File นั้นจะถูกเขียนใหม่ทุกครั้งที่มีการ เพิ่มข้อมูล ลบข้อมูล หรือแก้ไขข้อมูล จากระบบบริหารจัดการข้อมูล (Backend) ทั้งนี้ก็เพื่อประสิทธิภาพในการทำงานให้ดีที่สุด

4.2 ระบบส่วนติดต่อกับผู้ใช้งาน (Frontend) คือส่วนของการนำผลลัพธ์จากการวิเคราะห์ความรู้สึกนำมาแสดงผลบนเว็บไซต์ (Web Site) โดยส่วนนี้เปิดให้ผู้ใช้งานทั่วไปสามารถเข้ามาดูได้ โดยไม่จำเป็นต้องขอสิทธิ์ใด ๆ ในระบบส่วนติดต่อกับผู้ใช้งาน (Frontend) ใช้เทคนิกการเรียกใช้ข้อมูลโดยไม่จำเป็นต้อง Refresh หน้าเว็บเบราว์เซอร์ข้อมูลจะถูกส่งมาให้โดยอัตโนมัติผ่านทาง Socket IO

5. การทดสอบประสิทธิภาพของระบบ

5.1 การทดสอบประสิทธิภาพแบบไขว้ ระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษาบนเฟซบุ๊ก โดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ สำหรับภาษาไทย ในงานวิจัยนี้ใช้การทดสอบแบบไขว้ (K-fold cross-validation) เพื่อใช้ในการทดสอบประสิทธิภาพของโมเดลเนื่องจากผลที่ได้มีความน่าเชื่อถือ (เอกสิทธิ์, 2557) การวัด ประสิทธิภาพด้วยวิธี Cross-validation นี้จะทำการแบ่งข้อมูลออกเป็น K ส่วน แต่ละส่วนมีจำนวนข้อมูลเท่ากัน ในงานวิจัยชิ้นนี้กำหนดค่า k เท่ากับ 5 หมายถึงแต่ละชุดลองซึ่งมีจำนวนข้อมูล 12,548 ความคิดเห็นจะถูกแบ่งออกเป็น 5 ส่วนความคิดเห็นเท่า ๆ กัน หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของแบบจำลอง ทำวนไปเช่นนี้จนครบจำนวนที่แบ่งไว้ดังภาพ 3.26



ภาพที่ 3.26 การประเมินผลโดยวิธี K-fold Cross-validation

ที่มา : เอกสิทธิ์ (2557) “การแบ่งข้อมูลเพื่อทดสอบประสิทธิภาพของโมเดล” ค้นคืนวันที่ 10 ตุลาคม 2561 จาก http://dataminingtrend.com/2014/data-mining-techniques/cross-validation/

5.2 การประเมินผลและการทดสอบ เพื่อประเมินประสิทธิภาพการจำแนกข้อมูลในงานวิจัยนี้ จะใช้การวัดค่าความแม่นยำ (Precision) ค่าเรียกคืน (Recall) ค่าความถูกต้อง (Accuracy) และการวัดประสิทธิภาพโดยรวม (F-measure) เพื่อทดสอบความถูกต้อง โดยเริ่มพิจารณาค่าความถูกต้องจากตาราง Confusion Matrix ซึ่งเป็นการประเมินผลลัพธ์จากระบบเปรียบเทียบกับผลลัพธ์จริง ๆ ที่หาจากมนุษย์ ใช้วิธีคำนวณค่า Precision, Recall, Accuracy และ F-measure โดยมีวิธีการคำนวณค่าดังสมการที่ 2 ถึงสมการที่ 5

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Data\ Set} \quad (3)$$

$$F\text{-}measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

โดยที่ True Positive (TP) คือ จำนวนของความคิดเห็นที่เป็นเชิงบวกที่ระบบว่าเป็นเชิงบวก

True Negative (TN) คือ จำนวนความคิดเห็นที่เป็นเชิงลบที่ระบบว่าเป็นเชิงลบ

False Positive (FP) คือ จำนวนความคิดเห็นที่เป็นเชิงลบที่ระบบว่าเป็นเชิงบวก

False Negative (FN) คือ จำนวนความคิดเห็นเป็นเชิงบวกที่ระบบว่าเป็นเชิงลบ

Total Data Set คือ จำนวนของข้อมูลที่นำมาทดสอบทั้งหมด

บทที่ 4
ผลการดำเนินงาน

การศึกษาวิจัยเรื่องการพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ (Real-time) ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูล นาอีฟ เบย์ (Naive Bayes Classifier) สำหรับภาษาไทย มีวัตถุประสงค์เพื่อเป็นเครื่องมือช่วยในการวิเคราะห์ความรู้สึกที่นักศึกษามีต่อมหาวิทยาลัย เพื่อเป็นแนวทางในการนำไปปรับปรุงการจัดการเรียนการสอน และพัฒนาระบบการบริการของมหาวิทยาลัย ผู้วิจัยได้พัฒนาระบบเว็บไซต์ (Web Site) ให้ตรงกับความต้องการของผู้ใช้งานให้ได้มากที่สุด และได้ทำการทดสอบการทำงานของระบบเรียบร้อยแล้ว เพื่อหาข้อผิดพลาด พร้อมทั้งได้ทำการประเมินประสิทธิภาพการทำงานของระบบ โดยสามารถแบ่งผลการดำเนินการและการวิเคราะห์ข้อมูล เพื่อหาประสิทธิภาพของระบบงาน ดังต่อไปนี้

1. ผลการพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษาบนเฟซบุ๊ก โดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ สำหรับภาษาไทย

2. ผลการประเมินมาตรการความถูกต้องของค่าความแม่นยำและค่าเรียกคืนของระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ สำหรับภาษาไทย

1. ผลการพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษาบนเฟซบุ๊ก โดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ สำหรับภาษาไทย

ผู้วิจัยได้นำข้อความที่แสดงความคิดเห็นบนเฟซบุ๊กเพจมหาวิทยาลัยสุโขทัยธรรมาธิราช STOUniversity (https://fb.com/STOUniversity/) มาวิเคราะห์ความรู้สึก (Sentiment Analysis) แบบ 2 กลุ่ม คือ ความรู้สึกเชิงบวก (Positive) และความรู้สึกเชิงลบ (Negative) ในรูปแบบเว็บไซต์ (Web Site) โดยแบ่งการทำงานออกเป็น 2 ส่วน คือส่วนผู้ใช้งานระบบ และส่วนผู้ดูแลระบบ ซึ่งมีรายละเอียดดังนี้

1.1 ส่วนผู้ใช้งานระบบ เป็นส่วนที่ออกแบบและพัฒนาขึ้นมาเพื่อให้ผู้ใช้งานทั่วไปสามารถเข้ามาดูผลการวิเคราะห์ความรู้สึก แบบเรียลไทม์ (Real-time) โดยไม่จำเป็นต้องเป็นสมาชิกก็สามารถดูข้อมูลได้โดยความสามารถหลัก ๆ ของระบบในส่วนนี้มีดังต่อไปนี้คือ สามารถแสดง

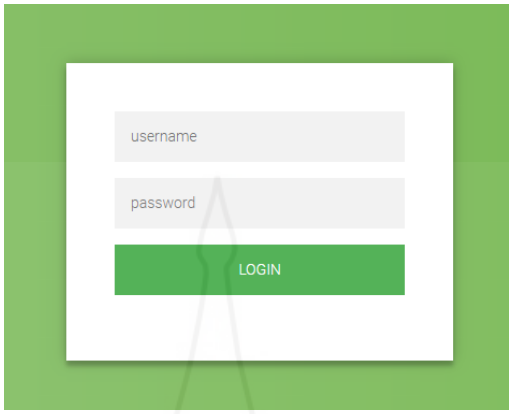
จำนวนผลการวิเคราะห์ความคิดเห็นเชิงบวกและจำนวนผลการวิเคราะห์ความรู้สึกที่เป็นเชิงลบได้ สามารถแสดงจำนวนผู้ที่กดติดตาม (Follow) และแสดงจำนวนผู้กดชื่นชอบเฟซบุ๊กเพจ (Like) ได้ สามารถแสดงผลการวิเคราะห์ข้อความที่เป็นความคิดเห็นเชิงลบพร้อมกับรายละเอียด และสามารถแสดงความคิดเห็นที่เป็นเชิงบวกได้ ซึ่งได้ผลดังภาพที่ 4.1

ภาพที่ 4.1 หน้าจอแสดงผลของระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ (Real-time)

1.2 ส่วนผู้ดูแลระบบ เป็นส่วนที่พัฒนาระบบขึ้นมาเพื่ออำนวยความสะดวกให้กับผู้ดูแลระบบ (Admin) สามารถบริหารจัดการคลังคำศัพท์ (Corpus) ดูรายงานสถิติต่าง ๆ อีกทั้งยังมีเครื่องมือทดสอบประโยคสามารถนำข้อความที่อยู่ในระบบมาทดสอบได้ทันที หรือแม้กระทั่งสามารถพิมพ์ข้อความเข้าไปทดสอบได้โดยตรง ระบบจะแสดงผลลัพธ์การตัดคำและวิเคราะห์ความรู้สึกให้ทันที การตรวจสอบสิทธิ์เข้าใช้งานระบบ รายงานความคิดเห็นทั้งในส่วนของความรู้สึกเชิงบวก ความรู้สึกเชิงลบหรือความรู้สึกที่เป็นกลาง รวมไปถึงการสมัครรับการแจ้งเตือนข้อความผ่านทางไลน์ (Line) โดยแบ่งออกเป็น 7 คุณลักษณะ (Feature) ซึ่งแต่ละคุณลักษณะมีรายละเอียดดังนี้

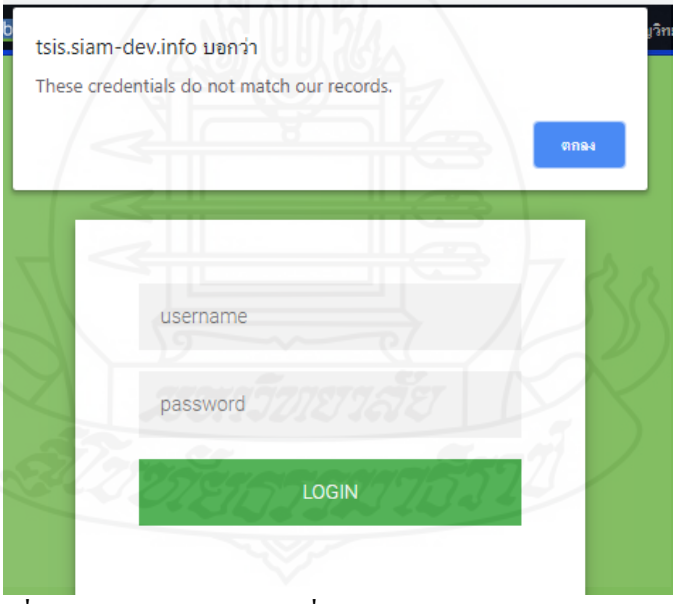
1.2.1 ระบบตรวจสอบการใช้งาน เป็นส่วนที่ออกแบบมาเพื่อควบคุมการเข้าถึงการใช้งานข้อมูลในระบบ โดยมีความสามารถหลัก ๆ อยู่ 3 ส่วนด้วยกันคือ 1) ตรวจสอบการเข้าใช้งาน 2) ตรวจสอบระยะเวลาหมดอายุในการใช้งานระบบ และ 3) สามารถให้ผู้ดูแลระบบออกจากระบบ

ได้ ส่วนที่ 1 ได้ออกแบบให้มีการกรอกบัญชีผู้ใช้งาน (Username) และรหัสผ่าน (Password) ดังภาพที่ 4.2



ภาพที่ 4.2 หน้าจอตรวจสอบผู้ใช้งาน

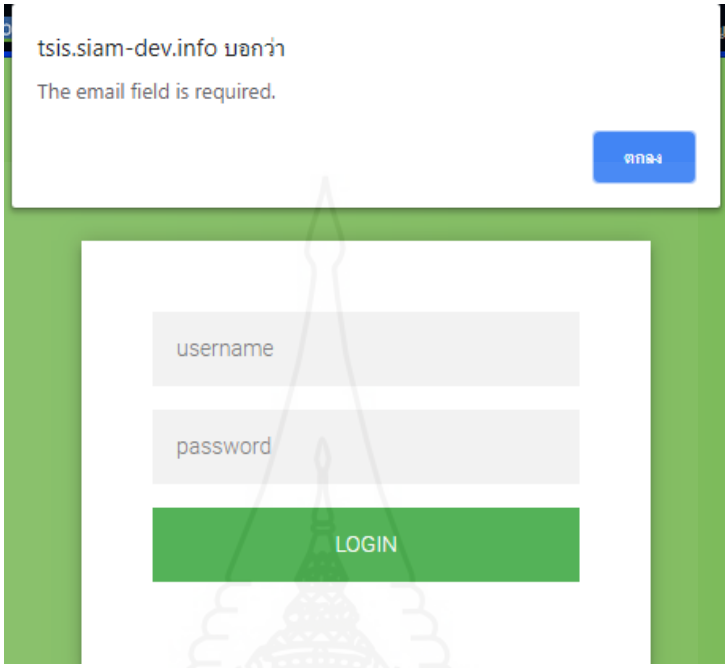
จากภาพที่ 4.3 เมื่อผู้ดูแลระบบกรอกบัญชีผู้ใช้งาน (Username) และรหัสผ่าน (Password) เสร็จแล้วทำการกดที่ปุ่ม LOGIN เพื่อส่งข้อมูลไปให้ระบบตรวจสอบกรณีที่ผู้ดูแลระบบกรณีที่ 1 กรอกข้อมูลผิดพลาดระบบจะแสดงผลดังภาพที่ 4.3



ภาพที่ 4.3 การแสดงข้อความเมื่อไม่พบข้อมูลผู้ดูแลระบบในระบบ

จากภาพที่ 4.3 กรณีแรกคือ ผู้ดูแลระบบกรอกข้อมูลครบทุกอย่างจากนั้น ผู้ดูแลระบบทำการส่งข้อมูลเข้าสู่ระบบ แล้วระบบทำการตรวจสอบพบว่า ข้อมูลบัญชีผู้ใช้งานหรือรหัสผ่านไม่ตรงกับฐานข้อมูลใด ๆ ในระบบ

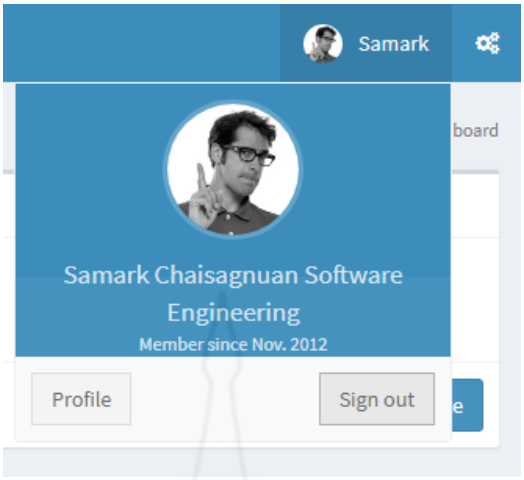
กรณีที่ 2 เมื่อผู้ดูแลระบบไม่ได้กรอกบัญชีผู้ใช้งาน (Username) ระบบจะแสดงข้อความแจ้งเตือนดังภาพที่ 4.4



ภาพที่ 4.4 ข้อความแจ้งเตือนกรณีไม่ได้กรอกบัญชีผู้ใช้งาน

ส่วนที่ 2 การตรวจสอบระยะเวลาการใช้งานในระบบในส่วนนี้จะเป็นส่วนที่ทำงานอยู่เบื้องหลัง คอยตรวจสอบว่าระยะเวลาในการเข้าสู่ระบบของผู้ดูแลระบบหมดอายุหรือยัง โดยเริ่มนับตั้งแต่วินาทีแรกที่ทำการเข้าสู่ระบบ ระยะเวลาที่ผู้ดูแลระบบสามารถอยู่ในระบบได้นั้นเป็นเวลา 24 ชั่วโมง หากระบบตรวจสอบแล้วพบว่า ระยะเวลาที่ใช้อยู่ในระบบนั้นหมดแล้ว ระบบจะนำพาไปยังหน้า Login ใหม่อีกครั้งหนึ่ง

ส่วนที่ 3 การออกจากระบบ เป็นส่วนที่ออกแบบไว้ให้ผู้ดูแลระบบที่ต้องการจะออกจากระบบโดยยังไม่ครบกำหนดเวลาหมดอายุ หรือป้องกันไม่ให้ผู้ใช้งานที่ไม่ใช่ผู้ดูแลระบบแอบเข้ามาใช้งานได้ ซึ่งได้ผลดังภาพที่ 4.5

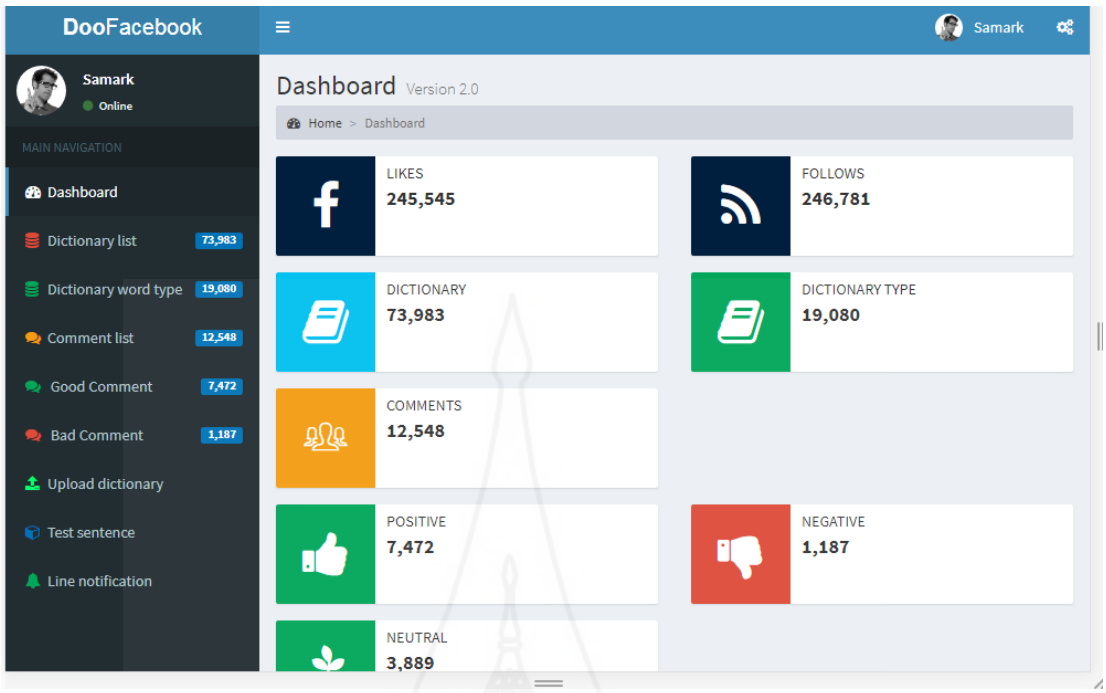


ภาพที่ 4.5 หน้าจอแสดงการออกจากระบบ

1.2.2 แผงควบคุม (Dashboard) เป็นส่วนที่ออกแบบและพัฒนาขึ้นมาเพื่อให้ผู้ดูแลระบบสามารถดูภาพรวมของข้อมูลแบบรวดเร็วได้ว่า ณ ขณะนี้มีจำนวนข้อมูลเท่าใด โดยมีคุณลักษณะ (Feature) อยู่ 7 อย่างด้วยกันคือ

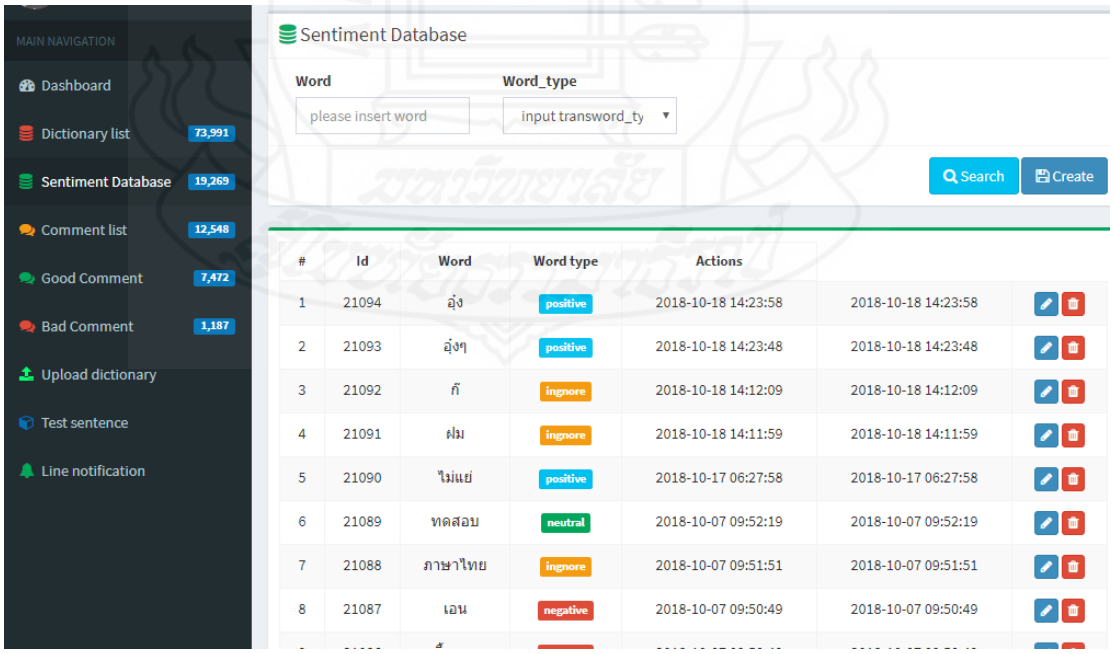
1) สามารถแสดงข้อมูลจำนวนผู้กดติดตามเฟซบุ๊กเพจได้
2) สามารถแสดงข้อมูลจำนวนผู้กดชื่นชอบเฟซบุ๊กเพจได้
3) สามารถแสดงจำนวนข้อมูลคลังคำศัพท์ได้
4) สามารถแสดงจำนวนข้อมูลคลังคำศัพท์ประเภทความรู้สึกได้
5) สามารถแสดงจำนวนผลการวิเคราะห์ความคิดเห็นเชิงบวกได้
6) สามารถแสดงจำนวนผลการวิเคราะห์ความคิดเห็นเชิงลบได้
7) สามารถแสดงจำนวนผลการวิเคราะห์ความคิดเห็นเป็นกลางได้

ซึ่งผลลัพธ์การพัฒนาระบบสามารถแสดงรายละเอียดได้ดังภาพที่ 4.6



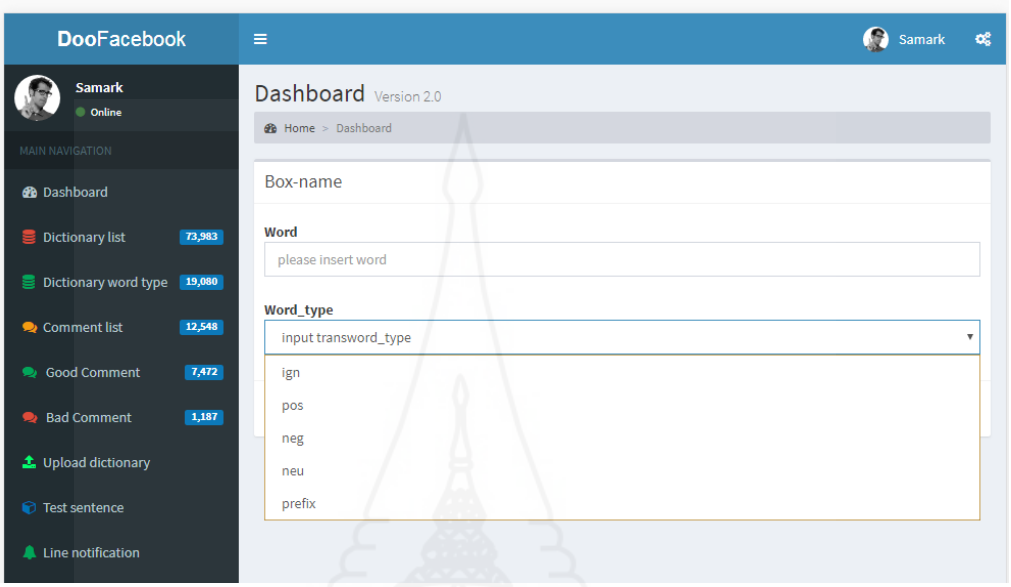
ภาพที่ 4.6 หน้าจอจัดการข้อมูลระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ (Real-time)

1.2.3 จัดการคำศัพท์ เป็นส่วนที่ออกแบบและพัฒนาขึ้นมาเพื่อให้ผู้ดูแลระบบสามารถบริหารจัดการคลังคำศัพท์ได้ง่าย สามารถทำการสืบค้นข้อมูลและยังรวมไปถึงการสร้างคำศัพท์ใหม่ แก้ไขข้อมูลคำศัพท์เดิม และลบข้อมูลคำศัพท์ที่ไม่ต้องการออกจากระบบ ได้ผลลัพธ์การพัฒนาระบบดังภาพที่ 4.7



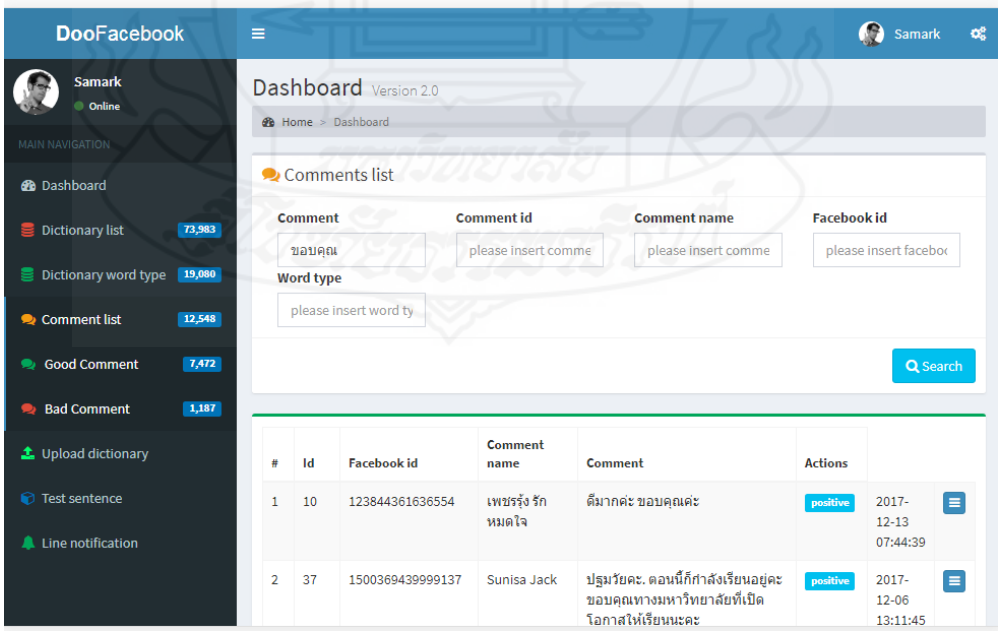
ภาพที่ 4.7 หน้าจอแสดงผลส่วนจัดการคำศัพท์ของระบบ

จากภาพที่ 4.7 เมื่อผู้ดูแลระบบมีความต้องการที่จะเพิ่มคำศัพท์ใหม่ แบบทีละคำ ผู้ดูแลระบบสามารถเพิ่มคำศัพท์และสถานะของคำศัพท์ใหม่เข้าไปในระบบได้ โดยกดที่ปุ่ม Create แล้วป้อนคำศัพท์พร้อมสถานะเข้าไปในระบบ ดังแสดงตัวอย่างในภาพที่ 4.8



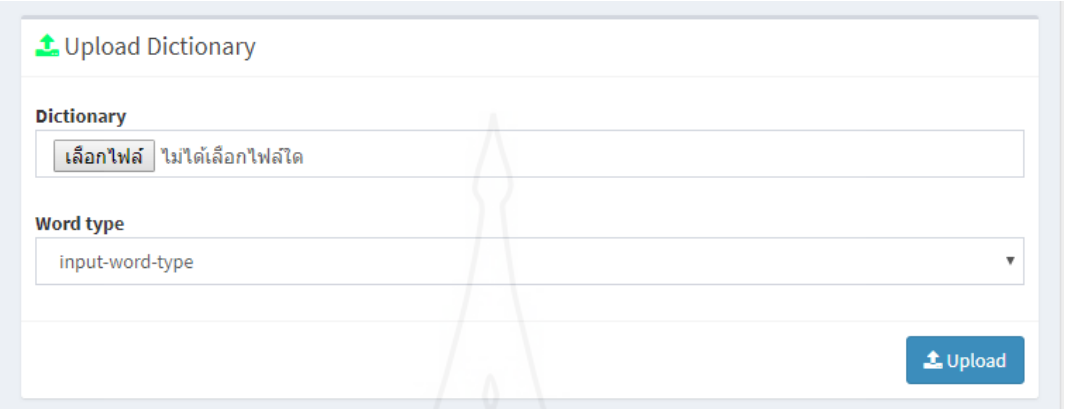
ภาพที่ 4.8 หน้าจอแสดงผลส่วนจัดการเพิ่มคำศัพท์ในพจนานุกรมของระบบ

1.2.4 รายงานความคิดเห็น (Comments list) เป็นส่วนรายงานข้อความผู้ใช้งานเฟซบุ๊ก โดยจะแสดงภาพรวมของข้อความทั้งหมด พร้อมทั้งสามารถค้นหาข้อความที่ต้องการดังภาพที่ 4.9



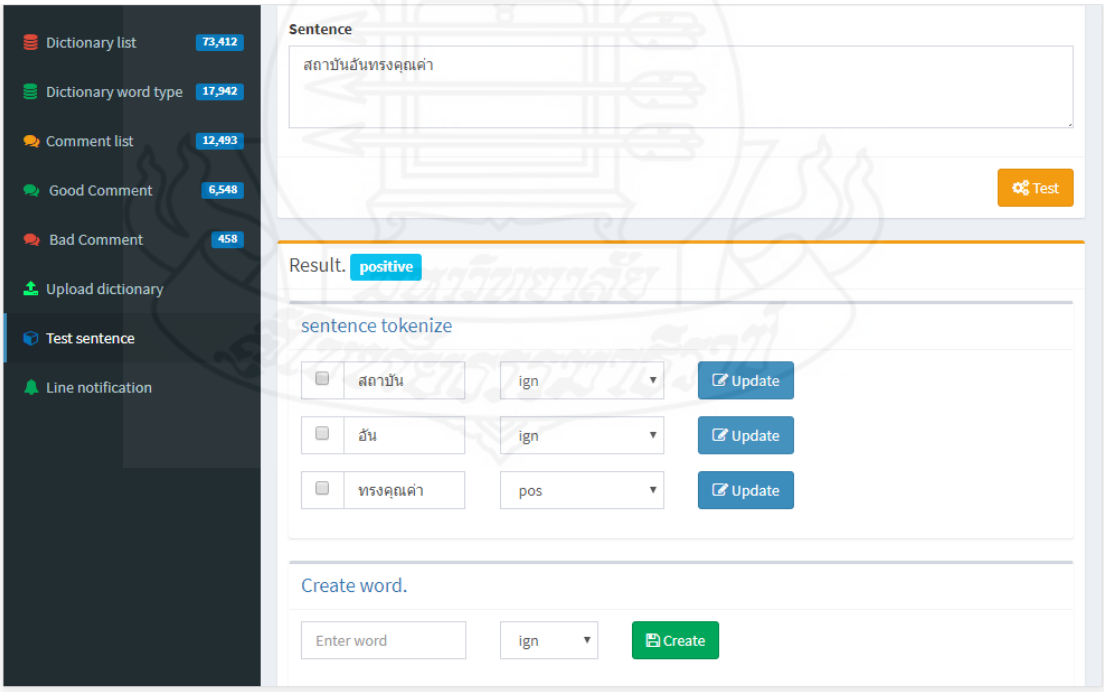
ภาพที่ 4.9 ส่วนรายงานข้อความผู้ใช้งานเฟซบุ๊ก

1.2.5 อัพโหลดคำศัพท์ (Upload Dictionary) เป็นส่วนอัพโหลดข้อมูลพจนานุกรมจำนวนมากเพิ่มเติมเข้ามาในระบบ โดยระบบออกแบบให้ผู้ดูแลระบบสามารถเลือกอัพโหลดได้จากไฟล์นามสกุล Text เพื่อเพิ่มประสิทธิภาพในการตัดคำได้ดียิ่งขึ้น ดังภาพที่ 4.10



ภาพที่ 4.10 หน้าต่างสำหรับอัพโหลดพจนานุกรม

1.2.6 เครื่องมือทดสอบประโยค (Sentence Test) เป็นส่วนสำหรับทดสอบประโยค โดยสามารถนำประโยคเข้าทดสอบการตัดคำและการแสดงสถานะของคำในประโยค เพื่อทดสอบความถูกต้อง พร้อมทั้งสามารถแก้ไขและเพิ่มสถานะของคำที่ตัดจากประโยคได้ ดังภาพที่ 4.11



ภาพที่ 4.11 หน้าจอสำหรับทดสอบประโยค

1.2.7 แจ้งเตือน Line Notification เป็นส่วนสำหรับส่งข้อความแจ้งเตือนไปยังไลน์ ในกรณีที่มีข้อความแสดงความคิดเห็นที่เป็นเชิงลบ โดยการเพิ่มบัญชีผู้ใช้สำหรับรับข้อความ ด้วยการแสกนคิวอาร์โค้ดสำหรับรับการแจ้งเตือน (QR Code) ผ่านทางแอปพลิเคชันไลน์ ดังภาพที่ 4.12

ภาพที่ 4.12 คิวอาร์โค้ดรับการแจ้งเตือน

จากภาพที่ 4.13 เมื่อผู้ดูแลระบบแสกนคิวอาร์โค้ดเพื่อรับการแจ้งเตือนแล้ว หากระบบพบข้อความที่มีความรู้สึกที่เป็นความคิดเห็นเชิงลบระบบจะทำการแจ้งเตือนผ่านทางไลน์ไปยังบัญชีผู้ใช้ของผู้ดูแลระบบพร้อมทั้งแสดงลิงค์ (Link) ที่ทำให้ผู้ดูแลระบบสามารถกดเข้าไปเพื่ออ่านรายละเอียดพร้อมทั้งสามารถตอบคำถามที่เฟซบุ๊กได้ทันที ดังภาพที่ 4.13

ภาพที่ 4.13 ข้อความแจ้งเตือน Line Notification

1.3 ผลการพัฒนาระบบตัดคำภาษาไทย

ในงานวิจัยชิ้นนี้ได้เลือกใช้กระบวนการตัดคำภาษาไทยด้วยวิธีการจับคู่ที่ยาวที่สุดและการจับคู่สูงสุด (Longest Matching and Maximal) โดยจะพิจารณาโดยการวนลูปจากซ้ายไปขวาตามหลักภาษาไทยเพื่อเทียบกับพจนานุกรม Lexicon – Thai สามารถอธิบายด้วยอัลกอริทึม (Algorithm) การตัดคำภาษาไทยโดยให้ผลลัพธ์การตัดคำดังตารางต่อไปนี้

ตารางที่ 4.1 ผลลัพธ์การตัดคำภาษาไทย

ตัวอย่างข้อความ	การตัดคำ
ภูมิใจมากที่ได้เรียนที่มสธ	ภูมิใจ|มาก|ที่|ได้|เรียน|ที่|มสธ
ดีใจที่ได้เป็นศิษย์มสธ	ดีใจ|ที่ได้เป็น|ศิษย์|มสธ
แชร์ได้ไหมครับท่านอาจารย์	แชร์|ได้ไหม|ครับ|ท่าน|อาจารย์
เปิดรับสมัครวันไหนครับ	เปิด|รับสมัคร|วัน|ไหนครับ
ยังไม่ได้รับหนังสือเลยค่ะ	ยัง|ไม่ได้รับหนังสือ|เลย|ค่ะ
ข้อสอบยากมาก	ข้อสอบ|ยากมาก

1.4 ผลการพัฒนาระบบวิเคราะห์ความรู้สึก

งานวิจัยชิ้นนี้ได้ใช้เทคนิคการจำแนกข้อมูลโดยใช้ตัวจำแนกนาอีฟ เบย์ (Naive Bayes Classifier) เพื่อวิเคราะห์ความรู้สึกโดยการให้น้ำหนักคำในประโยค ด้วยวิธีการเทียบคำที่ถูกตัดกับฐานข้อมูลประเภทของคำ ซึ่งให้ผลลัพธ์การวิเคราะห์ความรู้สึกดังตารางที่ 4.2

ตารางที่ 4.2 ผลลัพธ์การวิเคราะห์ความรู้สึก

ตัวอย่างข้อความ	การตัดคำ	ค่าคะแนน (pos/neg/neu)	ผลลัพธ์
ภูมิใจมากที่ได้เรียนที่มสธ	ภูมิใจ|มาก|ที่|ได้|เรียน|ที่|มสธ	**0.5000**/0.2500/0.2500	Positive
ดีใจที่ได้เป็นศิษย์มสธ	ดีใจ|ที่ได้เป็น|ศิษย์|มสธ	**0.4167**/0.3333/0.2500	Positive
แชร์ได้ไหมครับท่านอาจารย์	แชร์|ได้ไหม|ครับ|ท่าน|อาจารย์	0.2667/0.2667/**0.4667**	Neutral
เปิดรับสมัครวันไหนครับ	เปิด|รับสมัคร|วัน|ไหนครับ	0.2500/0.2500/**0.5000**	Neutral
ยังไม่ได้รับหนังสือเลยค่ะ	ยัง|ไม่ได้รับหนังสือ|เลย|ค่ะ	0.2500/**0.5000**/0.2500	Negative
ข้อสอบยากมาก	ข้อสอบ|ยากมาก	0.2500/**0.5000**/0.2500	Negative

ผลการศึกษาความรู้สึกของนักศึกษาบนเฟซบุ๊กโดยได้รวบรวมข้อมูลจากเฟซบุ๊กเพจของมหาวิทยาลัยสุโขทัยธรรมธิราช STOUniversity (https://fb.com/STOUniversity/) ผู้วิจัยได้เก็บรวบรวมข้อมูลบนเฟซบุ๊กเพจมหาวิทยาลัยสุโขทัยธรรมธิราชที่นำมาใช้ในระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ (Real-time) สำหรับภาษาไทย ซึ่งสามารถสรุปรายละเอียดของข้อมูลทั้งหมดได้จากตารางที่ 4.3 ดังนี้

ตารางที่ 4.3 สรุปข้อมูลที่ใช้งานระบบ

ข้อมูล	จำนวน	หน่วย
ชื่นชอบ (Like)	245,545	ผู้ใช้งานเฟซบุ๊ก
ติดตาม (Followed)	246,781	ผู้ใช้งานเฟซบุ๊ก
คลังคำศัพท์ (Corpus)	73,983	คำ
คลังคำศัพท์แยกประเภทความรู้สึก	19,081	คำ
ความคิดเห็น	12,548	ข้อความ
ความเห็นเชิงบวก (Positive)	7,472	ข้อความ
ความคิดเห็นเชิงลบ (Negative)	1,187	ข้อความ
ความคิดเห็นที่เป็นกลาง (Neutral)	3,889	ข้อความ

ผลการศึกษาความรู้สึกของนักศึกษาบนเฟซบุ๊กเพจมหาวิทยาลัยสุโขทัยธรรมาธิราช ด้วยวิธีการนำความคิดเห็นของนักศึกษาจำนวน 12,548 ความคิดเห็น บนเฟซบุ๊กเพจมหาวิทยาลัยสุโขทัยธรรมาธิราช มาวิเคราะห์บนระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ (Real-time) โดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ (Naive Bayes Classifier) สำหรับภาษาไทย ที่ผู้วิจัยได้พัฒนาขึ้นมา สามารถสรุปรายละเอียดของผลการวิเคราะห์ความรู้สึก โดยแบ่งออกเป็นความรู้สึกเชิงบวก ความรู้สึกเชิงลบ และคำถามที่พบบ่อย 5 อันดับแรกที่มีสถิติการพบคำศัพท์ในข้อความสูงสุดมีรายละเอียดดังตารางที่ 4.4 ถึง 4.6 ดังนี้

ตารางที่ 4.4 ผลการวิเคราะห์ความรู้สึกเชิงบวก (Positive) 5 อันดับแรก

ลำดับ	คำศัพท์	จำนวน	ตัวอย่างข้อความ
1	ยินดี	812	ยินดีที่ได้รู้จักเช่นกันค่ะขอให้ประสบความสำเร็จค่ะ
2	ภูมิใจ	514	ภูมิใจและดีใจที่จบรุ่นเดียวกันกลับคุณยายค่า
3	ขอบคุณ	391	ขอบคุณ มสธ.ที่ทำให้ผมจบ ป.ตรี ครับ
4	ดีใจ	142	ดีใจมากค่ะ สอบซ่อมผ่านแล้ว เหลืออีก5 วิชาก็จบแล้ว
5	รักมสธ	44	รักมสธ.และจะรักตลอดไปครับ

ตารางที่ 4.5 ผลการวิเคราะห์ความรู้สึกเชิงลบ (Negative) 5 อันดับแรก

ลำดับ	คำศัพท์	จำนวน	ตัวอย่างข้อความ
1	หนังสือ	128	หนังสือ ยังไม่ได้หนังสือเลยครับ ลงเทอม 1 จะเข้าเดือนธันวาคมแล้ว
2	อยาก	105	อยากต่อโทไม่มีเวลาทำงานยุ่งอายุมากขึ้นความจำไม่ค่อยดี
3	ยาก	53	สอบ 2 ครั้งนะ วิชานี้ แต่ยากสุดๆๆ คือเทคโนโลยี สอบ 3 ครั้งถึงผ่าน
4	กลัว	32	เกรดยังออกไม่ครบทุกตัวเลยคร้า กลัวลงไม่ทัน
5	ท้อ	24	เรียนจนท้อแล้ว วิชาสถิติ

ตารางที่ 4.6 ผลการวิเคราะห์คำถามที่พบบ่อย 5 อันดับแรก

ลำดับ	คำศัพท์	จำนวน	ตัวอย่างข้อความ
1	สมัคร	884	เปิดรับสมัครเมื่อไหร่ค่ะ
2	หนังสือ	166	อยากสอบถามว่าหนังสือส่งมาให้หรือยังค่ะ
3	เทียบโอน	118	จบป.ตรี บริหารธุรกิจ อยากเรียนศึกษาศาสตร์ มันเทียบโอนได้มั้ยคะ หรือต้องเรียนใหม่
4	ค่าใช้จ่าย	90	ค่าใช้จ่าย เอกสารการสมัครมีอะไรบ้างค่ะ แล้วชำระค่าใช้จ่ายอย่างไร
5	ผลสอบ	37	ประกาศผลสอบวันไหนค่ะ

2. ผลการประเมินมาตรการความถูกต้องของค่าความแม่นยำและค่าเรียกคืนของระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูลนาอีฟเบย์ สำหรับภาษาไทย

งานวิจัยมีการทดสอบประสิทธิภาพของระบบด้วยค่าความแม่นยำ (Precision) ค่าเรียกคืน (Recall) ค่าความถูกต้อง (Accuracy) และการวัดประสิทธิภาพโดยรวม (F-measure) เพื่อทดสอบความถูกต้องของข้อมูล ผู้วิจัยใช้วิธีการทดสอบโดยใช้โมเดล Cross Validation โดยการแบ่งข้อมูลออกเป็น 5 ชุดเท่ากัน และเก็บข้อมูล 1 ชุดไว้ทำการทดสอบ ทำซ้ำจนกว่าข้อมูลทุกชุดถูกนำมาทดสอบ ซึ่งเทคนิคนี้เรียกว่า 5-fold Cross Validation (เอกสิทธิ์,2557) ในงานวิจัยชิ้นนี้ได้นำข้อมูลความคิดเห็นของนักศึกษาบนเฟซบุ๊กเพจ มหาวิทยาลัยสุโขทัยธรรมาธิราชจำนวน 12,548 ข้อความ และเลือกส่วนที่สามารถนำมาทดสอบได้จำนวน 11,635 ข้อความ ซึ่งเป็นข้อความที่ไม่ใช่อักขระพิเศษหรือสัญลักษณ์ทางอารมณ์ (Emoji) แบ่งออกเป็น 5 ส่วนเท่า ๆ กัน โดยให้ข้อมูล 4 ส่วนเป็นข้อมูลชุดสำหรับฝึก (Training Data) และส่วนที่เหลือเป็นชุดสำหรับทดสอบ (Testing Data) จากนั้นทำการเปรียบเทียบผลการวิเคราะห์ระหว่าง ผลการวิเคราะห์ความรู้สึกด้วยระบบ และผลการวิเคราะห์ความรู้สึกจากผู้วิจัยได้ผลลัพธ์ดังต่อไปนี้

ตารางที่ 4.7 ผลการทดสอบประสิทธิภาพรอบที่ 1

Accuracy: 89.82%	**True Positive**	**True Neutral**	**True Negative**	**Class Precision**
Pred. Positive	1,165	45	5	95.88%
Pred. Neutral	41	788	2	94.83%
Pred. Negative	26	118	137	48.75%
Class Recall	**94.56%**	**82.86%**	**95.14%**	

ผลการทดสอบประสิทธิภาพรอบที่ 1 ได้ผลลัพธ์โดยแยกออกเป็น ผลการทดสอบค่าความถูกต้อง (Accuracy) 89.82% ค่าเรียกคืน (Recall) 90.85% ค่าความแม่นยำ (Precision) 79.82 % และค่าการวัดประสิทธิภาพโดยรวม (F-measure) 84.97 %

ตารางที่ 4.8 ผลการทดสอบประสิทธิภาพรอบที่ 2

Accuracy: 88.14%	**True Positive**	**True Neutral**	**True Negative**	**Class Precision**
Pred. Positive	1,118	8	30	96.71%
Pred. Neutral	55	130	150	38.81%
Pred. Negative	31	2	804	96.06%
Class Recall	**92.86%**	**92.86%**	**81.71%**	

ผลการทดสอบประสิทธิภาพรอบที่ 2 ได้ผลลัพธ์โดยแยกออกเป็น ผลการทดสอบค่าความถูกต้อง (Accuracy) 88.14% ค่าเรียกคืน (Recall) 89.14% ค่าความแม่นยำ (Precision) 77.19 % และค่าการวัดประสิทธิภาพโดยรวม (F-measure) 82.73 %

ตารางที่ 4.9 ผลการทดสอบประสิทธิภาพรอบที่ 3

Accuracy: 87.84%	**True Positive**	**True Neutral**	**True Negative**	**Class Precision**
Pred. Positive	949	68	3	93.04%
Pred. Neutral	37	989	8	95.65%
Pred. Negative	146	21	107	39.05%
Class Recall	**83.83%**	**91.74%**	**90.68%**	

ผลการทดสอบประสิทธิภาพรอบที่ 3 ได้ผลลัพธ์โดยแยกออกเป็น ผลการทดสอบค่าความถูกต้อง (Accuracy) 88.14% ค่าเรียกคืน (Recall) 88.75% ค่าความแม่นยำ (Precision) 75.91 % และค่าการวัดประสิทธิภาพโดยรวม (F-measure) 81.82 %

ตารางที่ 4.10 ผลการทดสอบประสิทธิภาพรอบที่ 4

Accuracy: 90.98%	**True Positive**	**True Neutral**	**True Negative**	**Class Precision**
Pred. Positive	838	67	2	92.39%
Pred. Neutral	22	1142	11	97.19%
Pred. Negative	83	27	136	55.28%
Class Recall	**88.87%**	**92.39%**	**91.28%**	

ผลการทดสอบประสิทธิภาพรอบที่ 4 ได้ผลลัพธ์โดยแยกออกเป็น ผลการทดสอบค่าความถูกต้อง (Accuracy) 90.98% ค่าเรียกคืน (Recall) 90.84% ค่าความแม่นยำ (Precision) 81.62 % และค่าการวัดประสิทธิภาพโดยรวม (F-measure) 85.98 %

ตารางที่ 4.11 ผลการทดสอบประสิทธิภาพรอบที่ 5

Accuracy: 91.35%	**True Positive**	**True Neutral**	**True Negative**	**Class Precision**
Pred. Positive	982	10	2	98.79%
Pred. Neutral	109	974	3	89.69%
Pred. Negative	16	61	166	68.31%
Class Recall	**88.71%**	**93.21%**	**97.08%**	

ผลการทดสอบประสิทธิภาพรอบที่ 5 ได้ผลลัพธ์โดยแยกออกเป็น ผลการทดสอบค่าความถูกต้อง (Accuracy) 91.35% ค่าเรียกคืน (Recall) 93.00% ค่าความแม่นยำ (Precision) 85.59 % และค่าการวัดประสิทธิภาพโดยรวม (F-measure) 89.14 %

ตารางที่ 4.12 สรุปผลการทดสอบประสิทธิภาพของระบบ

รอบที่	**ค่าความแม่นยำ (Precision)**	**ค่าเรียกคืน (Recall)**	**ค่าความถูกต้อง (Accuracy)**	**ค่าการวัดประสิทธิภาพโดยรวม (F-measure)**
1	79.82 %	90.85%	89.82%	84.97%
2	77.19 %	89.14%	88.14%	82.73%
3	75.91 %	88.75%	87.84%	81.22%
4	81.62 %	90.84%	90.98%	85.98 %
5	85.59 %	93.00%	91.35%	89.14 %
	80.03%	**90.52%**	**89.63%**	**84.81%**

จากตารางที่ 4.12 สามารถสรุปผลการทดสอบประสิทธิภาพของระบบได้ดังนี้ ผลการการประเมินค่าความแม่นยำ (Precision) 80.03% ผลการประเมินค่าเรียกคืน(Recall) 90.52% ผลการประเมินค่าความถูกต้อง 89.63% และผลการประเมินค่าการวัดประสิทธิภาพโดยรวม (F-measure) อยู่ที่ 84.81% ตามลำดับ

บทที่ 5
สรุปการวิจัย อภิปรายผล และข้อเสนอแนะ

จากการพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ (Real-time) ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ (Naive Bayes Classification) สำหรับภาษาไทย เพื่อใช้เป็นแนวทางในการนำไปปรับปรุงการจัดการเรียนการสอน และพัฒนาระบบการบริการของมหาวิทยาลัย สามารถสรุปผลการวิจัยได้ดังต่อไปนี้

1. สรุปการวิจัย
2. ปัญหาและอุปสรรค
3. ข้อเสนอแนะ

1. สรุปการวิจัย

ในปัจจุบันสื่อสังคมออนไลน์เช่นเฟซบุ๊กได้เข้ามาเป็นส่วนหนึ่งในชีวิตประจำวันของคนทั่วไป ความจำเป็นที่เจ้าของสินค้าผลิตภัณฑ์ หรือแม้กระทั่งหน่วยงานทั้งภาครัฐและเอกชน ที่ต้องจะต้องสร้างความสัมพันธ์ที่ดีให้กับผู้ใช้งานสื่อสังคมออนไลน์ ทั้งในส่วนที่เป็นลูกค้าหรือผู้ได้รับบริการ ทั้งนี้ก็เพื่อการสร้างภาพลักษณ์ที่ดี เพื่อเป็นการสร้างความเชื่อมั่นในตัวสินค้าหรือองค์กร การรับรู้ความรู้สึกจึงมีความจำเป็นอย่างยิ่ง ในกระบวนการสร้างภาพลักษณ์ที่ดี ซึ่งความคิดเห็นที่ดีก็ควรเก็บรักษาไว้เพื่อให้ใช้เป็นแนวทางปฏิบัติสืบต่อไป แต่ในส่วนของความรู้สึกที่ไม่ดีก็ควรนำมาพิจารณาปรับปรุงแก้ไข หรือชี้แจ้งทำให้เกิดความเข้าใจที่ชัดเจน ด้วยความที่ความคิดเห็นจากเฟซบุ๊กมีจำนวนข้อมูลที่มีปริมาณมาก การอ่านและตอบด้วยบุคคลากรจึงมีข้อจำกัดหลายอย่าง การนำระบบคอมพิวเตอร์เข้ามาช่วยในการวิเคราะห์นั้นจึงจำเป็นอย่างยิ่ง ที่จะช่วยแก้ปัญหาทั้งในเรื่องของการอ่าน ความคิดเห็นไม่ครบทุกข้อความ หรือตกหล่นไป

ดังนั้นผู้วิจัยได้นำเสนอการพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษาบนเฟซบุ๊ก โดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ สำหรับภาษาไทย ซึ่งประกอบไปด้วยสองส่วนหลักคือ ส่วนติดต่อกับผู้ใช้งานทั่วไป (Frontend) และส่วนบริหารจัดการข้อมูล (Backend) ในส่วนของผู้ใช้งานทั่วไปจะสามารถเข้ามาดูรายงานสถิติต่าง ๆ ผ่านทางเว็บไซต์ (Website) ที่ออนไลน์อยู่ตลอดเวลา อีกทั้งยังสามารถได้รับข้อมูลผลการวิเคราะห์แบบเรียลไทม์โดยไม่

จำเป็นต้องร้องขอข้อมูลไปยังเซิฟเวอร์ (Server) แต่เซิฟเวอร์ (Server) จะเป็นตัวส่งข้อมูลกลับมาที่เครื่องลูกข่าย (Client) แบบอัตโนมัติ และส่วนของผู้ดูแลระบบ (Backend) เป็นส่วนที่ออกแบบมาเพื่ออำนวยความสะดวกให้กับเจ้าหน้าที่ส่วนที่รับผิดชอบในการให้บริการโดยประกอบไปด้วย 1) ส่วนของการควบคุมการเข้าถึงข้อมูล 2) ส่วนของรายงานสถิติ 3) ส่วนของระบบการรายงานความคิดเห็น 4) ส่วนของการสมัครรับการแจ้งเตือนผ่านทาง Line 5) ส่วนของเครื่องมือทดสอบประโยค 6) ส่วนของรายงานความคิดเห็น และ 7) ส่วนของการบริหารจัดการคลังคำศัพท์ โดยทุกส่วนมีความเชื่อมโยงกันในแง่ ของการออกแบบระบบมาเพื่อใช้เป็นเครื่องมือในการวิเคราะห์ความรู้สึก ที่เกิดขึ้นบนเฟซบุ๊กเพจ ทั้งในส่วนที่ดีเพื่อนำมาใช้เป็นมาตรฐานการให้บริการ หรือในส่วนที่เป็นความคิดเห็นเชิงลบเพื่อติดตามแก้ปัญหา หรือชี้แจ้งทำความเข้าใจกับสิ่งที่เกิดขึ้นได้ทันท่วงที

ผู้วิจัยได้นำเสนอการพัฒนาเครื่องมือในการเพิ่มคำศัพท์ที่เกิดขึ้นใหม่ ด้วยการเพิ่มข้อมูลเข้าไปในฐานข้อมูลคลังคำศัพท์ ผ่านเครื่องมือที่ผู้วิจัยได้ออกแบบและพัฒนาขึ้นมา ให้สามารถใช้งานได้ง่ายไม่จำเป็นต้องมีความสามารถด้านการเขียนโปรแกรม ก็สามารถเพิ่มคำศัพท์ใหม่ได้ ทั้งนี้ยังช่วยให้ระบบมีประสิทธิภาพเพิ่มมากขึ้น

ผู้วิจัยได้นำเสนอเครื่องมือในการแจ้งเตือน กรณีที่ระบบวิเคราะห์ความรู้สึกออกมาเป็นความคิดเห็นเชิงลบ (Negative) จะมีการแจ้งเตือนข้อความไปยังผู้ดูแลระบบผ่านทาง Line ทั้งนี้ก็เพื่อประโยชน์ในการติดตามให้ช่วยแก้ไขปัญหาที่จะเกิดขึ้นให้ทันท่วงที โดยการส่งข้อความแจ้งเตือนพร้อมกับที่อยู่ต้นทาง (Link) แนบไปกับข้อความทำให้ผู้ดูแลระบบที่ สมัครรับการแจ้งเตือนสามารถรับข้อความแจ้งเตือนได้ทุกคน และสามารถกดเข้าไปยังที่อยู่ต้นทางเพื่ออ่านและตอบปัญหาที่เกิดขึ้นได้ สามารถสรุปผลได้ดังนี้

ได้ระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษาบนเฟซบุ๊ก โดยใช้ตัวจำแนกข้อมูลนาอีฟ เบย์ สำหรับภาษาไทย โดยแบ่งออกเป็นสองส่วนหลักคือ ส่วนติดต่อกับผู้ใช้งานทั่วไป (Frontend) และส่วนของผู้ดูและระบบ (Backend) สามารถทำงานและแสดงผลในรูปแบบ เว็บไซต์ออนไลน์ ผ่านระบบอินเตอร์เน็ต สามารถวิเคราะห์ข้อมูลแบบเรียลไทม์ และประมวลผลตั้งแต่กระบวนการรวบรวมข้อมูล กระบวนการตัดคำ กระบวนการวิเคราะห์ความรู้สึกโดยใช้ตัวจำแนกข้อมูลนาอีฟเบย์ (Naive Bayes Classification) รวมถึงกระบวนการแสดงผล สามารถทำงานได้แบบอัตโนมัติ

ผลการประเมินมาตรการความถูกต้องของค่าความแม่นยำและค่าเรียกคืนของระบบ ได้ดังนี้ ผลการประเมินค่าความแม่นยำ (Precision) 80.03% ผลการประเมินค่าเรียกคืน(Recall) 90.52% ผลการประเมินค่าความถูกต้อง 89.63% และผลการประเมินค่าการวัดประสิทธิภาพโดยรวม (F-measure) อยู่ที่ 84.81% ตามลำดับ

2. ปัญหาและอุปสรรค

2.1 การเข้าถึงข้อมูล จากปัญหาการพยายามจะแสวงหาประโยชน์จากเฟซบุ๊ก (Sunnywalker, 2018)โดยการนำข้อมูลผู้ใช้กว่า 50 ล้านบัญชี ไปเอื้อประโยชน์ในการหาเสียงเลือกตั้ง ทำให้เฟซบุ๊กเริ่มจำกัดการเข้าถึงข้อมูลของผู้ใช้ผ่าน API ทำให้อ่านข้อมูลได้น้อยลงจากเดิม โดยเฉพาะ "รายชื่อเพื่อน" ที่เป็นช่องโหว่ให้เกิดปัญหาข้อมูลรั่วไหล ทำให้ไม่สามารถเข้าถึงรายชื่อได้

2.2 Access Token เนื่องจากเฟซบุ๊กมีการปรับเรื่องอายุการใช้งาน Access Token เพื่อเข้าถึงข้อมูลที่อยู่ในเฟซบุ๊กเพจ (Facebook Page) โดยปรับลดลงจากเดิม ทำให้ต้องเพิ่มกระบวนการตรวจสอบอายุของ Access Token และกระบวนการขอ Access Token ใหม่

2.3 Application Review ในการเข้าไปใช้งานข้อมูลของเฟซบุ๊กโดยเฉพาะส่วนที่เป็นข้องมูลของเฟซบุ๊กเพจ นั้นจำเป็นจะต้องมีสิทธิ (Permission) ของเฟซบุ๊กชื่อว่าจัดการข้อมูลเพจ (Manage Page) ซึ่งทุกแอปพลิเคชัน (Application) ที่อยากจะใช้งานสิทธินี้ จะต้องผ่านกระบวนการแอปพลิเคชันรีวิว (Application Review) ของผู้ดูแลข้อมูลของเฟซบุ๊กก่อน ซึ่งในกระบวนการขออนุญาตนั้นไม่ได้หมายความว่าทุกแอปพลิคชันที่ขอไปจะผ่านเกณฑ์ กระบวนการหนึ่งในขั้นตอนการขอรีวิวแอปพลิเคชันคือการสร้างวีดีโอบันทึกกระบวนการใช้งานของแอปพลิเคชันที่ผู้ใช้งานจะเห็นตามความเป็นจริง กลายเป็นความลำบากอย่างหนึ่งในกรณีที่ยังไม่พัฒนาระบบขึ้นมาจริง ทำให้ไม่มีข้อมูลในส่วนนี้ไปนำเสนอให้เฟซบุ๊กได้ อีกขั้นตอนหนึ่งคือต้องเขียนอธิบายให้ผู้ดูแลข้อมูลเฟซบุ๊กเข้าใจว่าเหตุผลอะไรที่จำเป็นต้องใช้สิทธิในการจัดการเฟซบุ๊กเพจ ถ้าเขียนอธิบายไม่ดีพอ จะกลายเป็นเหตุผลให้แอปพลิเคชันนั้นก็จะไม่ผ่านเกณฑ์ ยิ่งไปกว่านั้นจากสถาณะการณ์ข้อมูลของผู้ใช้งานเฟซบุ๊ก รั่วไหล ทำให้กระบวนการขอสิทธิ์ มีความเข้มงวดมากยิ่งขึ้นหลายเท่า

2.4 การใช้งานข้อมูลในเฟซบุ๊กเพจ (Facebook Page) นั้น เฟซบุ๊ก (Facebook) อนุญาตให้เฉพาะผู้ดูแลเฟซบุ๊กเพจ เท่านั้นที่ทำได้ กรณีที่เป็นแค่นักพัฒนาระบบ (Developer) จะได้เฉพาะข้อมูลเพียงบางอย่างที่อนุญาตเท่านั้นเช่น ความคิดเห็นของผู้ติดตาม แต่ไม่ได้ชื่อผู้ที่แสดงความคิดเห็น ซึ่งเป็นการปรับวิธีการเข้าถึงข้อมูลของเฟซบุ๊ก เพื่อเพิ่มมาตรการด้านความปลอดภัยของข้อมูล งานวิจัยชิ้นนี้ได้พัฒนาระหว่างช่วงก่อนที่เฟซบุ๊ก จะออกนโยบายการเข้าถึงข้อมูล จึงสามารถได้ข้อมูลบางส่วนที่ครบถ้วนไม่มีการปิดกั้นใด ๆ แต่หลังจากเดือน เมษายน 25561 ที่ผ่านมากระบวนการดึงข้อมูลมาใช้งานทำได้เพียงแค่ดึงข้อมูลได้เฉพาะความคิดเห็นและไม่ได้ชื่อที่ใช้งานมาด้วย

3. ข้อเสนอแนะ

3.1 การทำงานของระบบจะมีประสิทธิภาพดีที่สุดในหมวดหมู่ของเฟซบุ๊กเพจที่เกี่ยวข้องกับการศึกษา ซึ่งหากต้องการนำไปใช้ในหมวดหมู่อื่น ๆ ควรนำเข้าข้อมูลคลังคำศัพท์ที่แยกประเภทความรู้สึกตามหมวดหมู่นั้น ๆ เพื่อเพิ่มประสิทธิภาพให้ระบบ

3.2 เฟซบุ๊กมีการปรับนโยบายการเข้าใช้งานข้อมูลผ่าน Facebook Graph API ในปีที่ผ่านมาด้วยกันหลายครั้ง และนโยบายล่าสุดหากจะนำระบบไปใช้งานงานต่อจำเป็นต้องขออนุญาตการใช้งานข้อมูลไปทางเฟซบุ๊ก และต้องเป็นในลักษณะของการดำเนินการทางธุรกิจเท่านั้น โดยจะต้องใช้หลักฐานฐานเช่นหนังสือประกอบธุรกิจพาณิชย์ที่ทางราชการออกให้ หนังสือการเสียภาษีอากรเป็นต้น

3.3 ระบบนี้สามารถทำงานได้ดีบนภาษาไทยเท่านั้น ซึ่งสามารถนำไปพัฒนาต่อยอดให้รองรับภาษาอื่น ๆ ที่ต้องการได้ ด้วยการพัฒนาฐานข้อมูลคลังคำศัพท์ของภาษานั้น ๆ ขึ้นมา

3.4 สามารถนำระบบนี้ไปพัฒนาต่อยอดไปเป็นระบบอื่น ๆ ได้เช่นระบบตอบกลับความคิดเห็นอัตโนมัติ (Chat Bot) ซึ่งจะช่วยลดภาระงานให้เจ้าหน้าที่ ได้เป็นอย่างดี

บรรณานุกรม

บรรณานุกรม

กานดา รุณนะพงศา. (2549). *การตัดคำภาษาไทยด้วยวิธีปรับปรุงกฎและพจนานุกรมแบบใหม่.* คณะวิศวกรรมศาสตร์ สาขาวิชาวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น, ขอนแก่น.

คลังข้อมูลภาษาไทยแห่งชาติ ในพระราชูปถัมภ์สมเด็จพระเทพรัตนราชสุดาฯ สยามบรมราชกุมารี. (2561) ฐานข้อมูลคลังคำศัพท์. ค้นคืนวันที่ 10 สิงหาคม 2561 จาก http://www.arts.chula.ac.th/ling/tnc/?smd_process_download=1&download_id=144

นภดล สิทธิเลิศ. (2556). *การพยากรณ์อารมณ์ความรู้สึกจากการโพสต์ข้อความบนเฟซบุ๊ก (Facebook).*(สารนิพนธ์วิทยาศาสตรมหาบัณฑิต) มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ.

นริศร์ พรหมบุตร. (2550). *การทำไมน์นิ่งความคิดเห็นสินค้า กรณีศึกษาโทรศัพท์มือถือ.*(สารนิพนธ์วิทยาศาสตรมหาบัณฑิต) สถาบันเทคโนโลยีพระจอม เกล้าพระนครเหนือ, กรุงเทพฯ.

บุญเสริม กิจศิริกุล .(2546). *อังกอริทึมการทำเหมืองข้อมูล.* รายงานวิจัยฉบับสมบูรณ์ โครงการวิจัยร่วมภาครัฐและเอกชน ปีงบประมาณ 2545 ภาควิชา วิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย.

บ้านจอมยุทธ. (2543). *ระบบเรียลไทม์.* ค้นคืนวันที่ 10 สิงหาคม 2561 จาก https://www.baanjomyut.com/library_2/extension-1/internet/10.html

ประเสริฐ ศรีราชพัฒน์ (2555). *หน่วยเสียงสระ.* ค้นคืนวันที่ 10 สิงหาคม 2561 จาก http://www.thaigoodview.com/node/144890

ปรีดาวรรณ เกษเมธีการุณ. (2559). *การวิเคราะห์ความรู้สึกจากวีดีทัศน์บทวิจารณ์ภาษาไทยโดยใช้เทคนิคการทำเหมืองข้อมูล.*(ดุษฎีนิพนธ์ ปรัชญาดุษฎีบัณฑิต) มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ.

พนิดา ทรงรัมย์. (2559). การจำแนกความคิดเห็นทางการเมืองบนเครือข่ายสังคมออนไลน์ โดยใช้วิธีการจำแนกแบบสัมพันธ์. *วารสารวิทยาศาสตร์และเทคโนโลยี มทร.ธัญบุรี.* Vol.6 No.1, pp. 83-93.

สมปรารถนา รัทยานนท์. (2535). *โครงสร้างข้อมูลสำหรับพจนานุกรมอิเล็กทรอนิกส์ภาษาไทย.* (วิทยานิพนธ์ปริญญามหาบัณฑิต) บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.

วิรัชศร เลิศล้ำ วาณิช. (2536). *การตัดคำภาษาไทยในระบบแปลภาษา.* กรุงเทพมหานคร : ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ. 50-55.

วฤษาย์ ร่มสายหยุด. (2561). *หลักการวิศวกรรมซอฟต์แวร์.* นนทบุรี : มหาวิทยาลัยสุโขทัยธรรมาธิราช.

วิกิพีเดีย สารานุกรมเสรี. (2561). *เฟซบุ๊ก.* ค้นคืนวันที่ 10 สิงหาคม 2561 จาก https://th.wikipedia.org/wiki/เฟซบุ๊ก.

วิกิพีเดีย สารานุกรมเสรี. (2561). *ไลน์* (โปรแกรมประยุกต์). ค้นคืนวันที่ 10 สิงหาคม 2561 จาก https://th.wikipedia.org/wiki/ไลน์_(โปรแกรมประยุกต์).

ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ. (2547). *ฐานข้อมูลพจนานุกรมภาษาไทย.* http://lexitron.nectec.or.th/dnload_data/lexitron-data.zip.

เอกสิทธิ์. (2557). *การแบ่งข้อมูลเพื่อทดสอบประสิทธิภาพของโมเดล.* ค้นคืนวันที่ 10 ตุลาคม 2561 จาก http://dataminingtrend.com/2014/data-mining-techniques/cross-validation/.

Saixiii. (2560). *Line API* ค้นคืนวันที่ 10 ตุลาคม 2561 จาก https://saixiii.com/chapter2-line-api-official/.

wp. (2561). สถิติผู้ใช้ดิจิทัลทั่วโลก “ไทย” เสพติดเน็ตมากสุดในโลก-“กรุงเทพ” เมืองผู้ใช้ Facebook. สูงสุดค้นคืนวันที่ 7 กรกฎาคม 2561 จาก https://www.brandbuffet.in.th/2018/02/global-and-thailand-digital-report-2018/.

Choochart Haruechaiyasak*, Alisa Kongthon*,Pornpimon Palingoon*, and Kanokorn Trakultaweekoon* . (2018). S-Sense: A Sentiment Analysis Framework for Social Media Monitoring Applications.Information Technology Journal.

C.*Troussas,*M.*Virvou,*et al. (2013). Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning.

*Fourth International Conference on Information, Intelligence, Systems and Applications (IISA). 41-6.

Daniel Jurafsky and James H. Martin. (2018). Naive Bayes And Sentiment Classification, Speech and Language Processing. 61-79.

Muhammad Awis Jamaluddin Johari, Amru Yusrin Amruddin and Dr. Chew Yew Choong. (2015). SYSTEM TO AUTOMATICALLY CALCULATE THE ACCURACY (SACA) . Software Development Lab, ICT Department, MIMOS Berhad, Malaysia.

Tetsuya and Jeonghee. (2003). Sentiment Analysis Capturing favorability using Natural Language Processing. IBM Research, Tokyo Research Laboratory. 1623-14 Shimotsuruma, Yamato-shi, Kanagawa-ken, 242-8502, Japan.

Vishwasa Navada K, Naveen D Chandavarkar. (2015). Real time Extraction of Sentiment and Analysis for Decision Making. Dept. of Computer Science, NMAM Institute of Technology Nitte, Karnataka, India.

Wunnasri,*T.*Theeramunkong*and*C.*Haruechaiyasak. (2013). Solving Unbalanced Data for Thai Sentiment Analysis. *International Joint Conference on Computer Science and Software Engineering (JCSSE). 200-205.

ประวัติผู้วิจัย

ชื่อ	นายสมัคร ชัยสงวน
วัน เดือน ปีเกิด	6 กุมภาพันธ์ 2532
สถานที่เกิด	อำเภอสว่างแดนดิน จังหวัดสกลนคร
ประวัติการศึกษา	สาขาวิชาวิทยาศาสตร์และเทคโนโลยี แขนงเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยสุโขทัยธรรมาธิราช พ.ศ 2560
สถานที่ทำงาน	บริษัททรู คอร์ปอเรชั่นจำกัด มหาชน แขวงดินแดง เขตดินแดง กรุงเทพมหานคร 10400
ตำแหน่ง	นักเขียนโปรแกรมอาวุโส (Senior Programmer)