การประมวลผลภาษาธรรมชาติจากข้อมูลสื่อสังคมออนไลน์เชิงพื้นที่เพื่อประมาณการเกิดอุบัติเหตุ

NATURAL LANGUAGE PROCESSING TECHNIQUESWITH SOCIAL MEDIA SPATIAL DATA FOR ACCIDENT ESTIMATION

สิทธิพัฒน์ เปี่ยมใจสว่าง

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

2566

การประมวลผลภาษาธรรมชาติจากข้อมูลสื่อสังคมออนไลน์เชิงพื้นที่เพื่อประมาณการเกิดอุบัติเหตุ

สิทธิพัฒน์ เปี่ยมใจสว่าง

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิชาภูมิสารสนเทศ
คณะสังคมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2566
ลิขสิทธิ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ

NATURAL LANGUAGE PROCESSING TECHNIQUESWITH SOCIAL MEDIA SPATIAL DATA FOR ACCIDENT ESTIMATION

SITTHIPAT PIEMJAISWANG

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of MASTER OF SCIENCE

(Geoinformatics)

Faculty of Social Sciences, Srinakharinwirot University

2023

Copyright of Srinakharinwirot University

ปริญญานิพนธ์

เรื่อง

การประมวลผลภาษาธรรมชาติจากข้อมูลสื่อสังคมออนไลน์เชิงพื้นที่เพื่อประมาณการเกิดอุบัติเหตุ

ของ

สิทธิพัฒน์ เปี่ยมใจสว่าง

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาภูมิสารสนเทศ

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)

คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบปากเปล่าปริญญานิพนธ์

..... ที่ปรึกษาหลัก

(ผู้ช่วยศาสตราจารย์ ดร.อสมาภรณ์ สิทธิ)

..... ประธาน

(ผู้ช่วยศาสตราจารย์ ดร.ณรงค์ พลีรักษ์)

..... ที่ปรึกษาร่วม

(ผู้ช่วยศาสตราจารย์ ดร.ปกรณ์ เมฆแสงสวย)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ชูเดช โลศิริ)

ชื่อเรื่อง	การประมวลผลภาษาธรรมชาติจากข้อมูลสื่อสังคมออนไลน์เชิงพื้นที่เพื่อประมาณการเกิดอุบัติเหตุ
ผู้วิจัย	สิทธิพัฒน์ เปี่ยมใจสว่าง
ปริญญา	วิทยาศาสตรมหาบัณฑิต
ปีการศึกษา	2566
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. อสมาภรณ์ สิทธิ
อาจารย์ที่ปรึกษาร่วม	ผู้ช่วยศาสตราจารย์ ดร. ปกรณ์ เมฆแสงสวย

การศึกษาการประมวลผลภาษาธรรมชาติในการประมาณการเกิดอุบัติเหตุจากสื่อสังคมออนไลน์มีวัตถุประสงค์เพื่อการจัดทำฐานข้อมูลเชิงพื้นที่จากการรายงานอุบัติเหตุผ่านสื่อสังคมออนไลน์ มาจากการที่ผู้ใช้แบ่งปันข้อมูลรูปภาพหรือแสดงความคิดเห็นต่างๆ รวมถึงการอ้างถึงที่มีความสำคัญกับอุบัติเหตุ โดยพิจารณาการเกิดอุบัติเหตุเป็น 3 ช่วงเวลาคือ ช่วงเช้า ช่วงกลางวัน และช่วงเย็น งานวิจัยนี้นำเทคนิคเหมืองดัชนีข้อความโดยการนำข้อความที่สกัดได้จากสื่อสังคมออนไลน์มาหาความสัมพันธ์ของข้อมูล การดำเนินงานวิจัยแบ่งออกเป็น 3 ขั้นตอนหลัก คือ 1.การสร้างฐานข้อมูลเกี่ยวกับอุบัติเหตุที่ได้จากทวิตเตอร์ 2.การประมวลผลข้อความของและการหาความสัมพันธ์ของข้อมูลอุบัติเหตุ และ 3.กระบวนการเข้ารหัสจากข้อความให้เป็นข้อมูลเชิงพื้นที่ (Geocoding) ผลการวิจัยพบว่า การหากฎของความสัมพันธ์ระหว่างข้อความและความหนาแน่นเชิงพื้นที่ของการเกิดอุบัติเหตุจากสื่อสังคมออนไลน์ และเปรียบเทียบพื้นที่การเกิดอุบัติเหตุเพื่อหาปัจจัยในการเกิดโดยในแต่ละแบ่งช่วงเวลาได้แก่ เช้า กลางวัน เย็น มีความเชื่อมโยงกับลักษณะการใช้งานสื่อสังคมออนไลน์และมีความสอดคล้องกับสถิติการเกิดอุบัติเหตุ ที่มีการรายงานสภาพการจราจรและการรายงานอุบัติเหตุในแต่ละช่วงเวลา ผลการศึกษาที่ได้จากการประมวลภาษาธรรมชาติและการหากฎของความสัมพันธ์ดัชนีข้อความที่ได้จากทวิตเตอร์ สามารถอธิบายถึงลักษณะของการรายงานอุบัติเหตุจากสื่อสังคมออนไลน์ และช่วยคาดการณ์เหตุการณ์ที่เกิดขึ้นโดยอธิบายถึงการกระจายเชิงพื้นที่ของการระบุตำแหน่งของอุบัติเหตุ ซึ่งสามารถนำไปใช้ประโยชน์ในการวางแผนและพัฒนาโครงข่ายถนน การเสริมสร้างความปลอดภัยบนท้องถนน การเข้าถึงพื้นที่ โครงสร้างพื้นฐานการคมนาคมขนส่งที่มีความสำคัญต่อเหตุการณ์นั้นๆ ได้อย่างมีประสิทธิภาพมากขึ้น

คำสำคัญ : การประมวลผลภาษาธรรมชาติ,การระบุตำแหน่งทางภูมิศาสตร์,การคาดการณ์อุบัติเหตุ

Title	NATURAL LANGUAGE PROCESSING TECHNIQUESWITH SOCIAL MEDIA SPATIAL DATA FOR ACCIDENT ESTIMATION
Author	SITTHIPAT PIEMJAISWANG
Degree	MASTER OF SCIENCE
Academic Year	2023
Thesis Advisor	Assistant Professor Dr. Asamaporn Sitthi
Co Advisor	Assistant Professor Dr. Pakorn Meksangsouy

This study employs Natural Language Processing (NLP) techniques to estimate accident occurrences by analyzing user-generated content on social media platforms. The primary objective is to construct a spatial database of accidents by extracting relevant information, including the pictures and comments shared on social media. The study specifically focused on ascertaining the mentioned locations or areas of accidents during the morning, afternoon, and evening. Text mining techniques were applied to process accident data extracted from social media, identifying the frequency of accident-related words, classifying accident types, and establishing their relations. The research methodology was comprised of three key steps. Firstly, querying accident data using the Twitter API. Secondly, conducting text mining on accident events and utilizing association rule mining techniques for accident data. Thirdly, employing geocoding to estimate the relative location of accident occurrences and determine the spatial density of accidents. The findings of the study highlighted the identification of rules governing the relationship between text and the spatial density of accidents reported on social media. A comparison of accident-prone areas during distinct time periods, namely morning, afternoon, and evening, revealed connections with the usage patterns of social media. The results obtained from natural language processing and the elucidation of rules governing text indices from Twitter shed light on the characteristics of accident reports on social media. This understanding predicts future events by explaining the spatial distribution of accident locations. These insights can be instrumental in planning and developing road networks, enhancing road safety measures, and improving efficiency in accessing areas, infrastructure, and transportation crucial to specific events.

Keyword : Natural language processing Geocoding Accident estimation

กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้สำเร็จได้ด้วยดี ทั้งนี้เพราะได้รับความอนุเคราะห์และเมตตาจากหลายท่านผู้มีพระคุณผู้วิจัยขอขอบพระคุณทุกท่านเป็นอย่างยิ่ง

ขอขอบคุณประธานกรรมการปริญญานิพนธ์ ผศ.ดร.ณรงค์ พลีรักษ์ อาจารย์ที่ปรึกษาปริญญานิพนธ์หลัก ผศ.ดร.อสมาภรณ์ สิทธิ และอาจารย์ที่ปรึกษาปริญญานิพนธ์ร่วม ผศ.ดร.ปกรณ์ เมฆแสงสวย กรรมการปริญญานิพนธ์ ผศ.ดร.ชูเดช โลศิริ ที่ให้ความเมตตาและกรุณาให้คำปรึกษาทางด้านวิชาการ แนะแนวทางการดำเนินงาน และการแก้ปัญหาต่างๆ อันเป็นประโยชน์อย่างยิ่งในการทำงานวิจัยนี้ ตลอดจนการตรวจสอบการทำงานวิจัย อีกทั้งช่วยแนะสิ่งที่ขาดตกบกพร่องตลอดจนในกำลังใจในการทำงานวิจัยนี้ให้สำเร็จลุล่วง

ขอกราบขอบพระคุณอาจารย์ทุกท่าน ที่มอบความรู้ให้แก่ผู้วิจัย ในสิ่งที่เป็นประโยชน์อย่างยิ่ง และขอบพระคุณ พี่ๆ เพื่อนๆ น้องๆ รวมถึงเจ้าหน้าที่ภาควิชาภูมิศาสตร์ คณะสังคมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ที่ให้คำปรึกษาและช่วยเหลือตลอดช่วงระยะเวลาการศึกษา

ผู้วิจัยรู้สึกซาบซึ้งในกำลังใจและความเมตตาช่วยเหลือจากทุกท่าน จึงขอขอบพระคุณอย่างสูงมา ณ โอกาสนี้

สิทธิพัฒน์ เปี่ยมใจสว่าง

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฎ

สารบัญรูปภาพ ..... ฎ

บทที่1 บทนำ 1

1. ที่มาและความสำคัญ 1

2. วัตถุประสงค์ของงานวิจัย 2

3. ความมุ่งหมายของงานวิจัย 3

4. ขอบเขตของงานวิจัย 3

5. คำสำคัญ 4

6. นิยามศัพท์เฉพาะ 4

7. ประโยชน์ที่คาดว่าจะได้รับ 6

8. กรอบแนวคิดงานวิจัย 6

บทที่2 เอกสารและงานวิจัยที่เกี่ยวข้อง 7

1 ข้อมูลทั่วไปของพื้นที่ศึกษา 7

1.1 สภาพภูมิประเทศ 9

1.2 สภาพภูมิอากาศ 9

1.3 การใช้ประโยชน์ที่ดินจังหวัดกรุงเทพมหานคร 9

1.4 ระบบคมนาคมขนส่งทางบกในกรุงเทพมหานคร 10

1.5 ถนนที่มีการจราจรหนาแน่น ในจังหวัดกรุงเทพมหานคร 12
2. สื่อสังคมออนไลน์ (Social Media) 14
3 การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) 16
3.1 การประมวลผลข้อความ (Text Processing) 18
3.1.1 การเตรียมข้อมูล (Text Preprocessing phase) 19
3.1.2 การสกัดข้อมูล (Information Extraction) และการกรองข้อมูล (Data Filtration) 20
3.1.3 การทำกลุ่มข้อมูล (Word Cloud) และความถี่ข้อมูล (Frequency) 20
3.2 การหาความสัมพันธ์ของเหมืองดัชนีถ้อยคำ 21
3.2.1 การจัดกลุ่มข้อมูล (Data Clustering) 21
3.2.2 กฎของการหาความสัมพันธ์ (Association Rule Mining: ARM) 22
3.3 อัลกอริทึมอพริโอริ (Apriori Algorithm) 23
4. ระบบสารสนเทศภูมิศาสตร์ (Geographic Information System: GIS) 25
4.1 ข้อมูลในระบบสารสนเทศภูมิศาสตร์ 26
4.2 องค์ประกอบของระบบสารสนเทศภูมิศาสตร์ 27
4.2.1 ฮาร์ดแวร์ (Hardware) 27
4.2.2 ซอฟท์แวร์ (Software) 27
4.2.3 วิธีการหรือขั้นตอนการปบัติงาน (Methodology) 28
4.2.4 การนำเข้าข้อมูล (Data Input) 28
4.2.5 การจัดการข้อมูล (Data Management) 28
4.2.6 การวิเคราะห์ข้อมูล (Data Analysis) 28
4.2.7 การแสดงผล (Data Display) 28
4.2.8 บุคลากร (People) 29

4.2.9 ข้อมูล (Data) 29
4.3 ระบบฐานข้อมูล (Database) 29
4.4 กระบวนการแปลงข้อความเป็นพิกัดทางภูมิศาสตร์ (Geocoding) 30
5. เอกสารและงานวิจัยที่เกี่ยวข้อง 33
บทที่3 วิธีดำเนินงานวิจัย 37
1. แผนการดำเนินงาน 37
2. อุปกรณ์และเครื่องมือที่ใช้ในงานวิจัย 38
2.1 คุณสมบัติเครื่องคอมพิวเตอร์ที่ใช้ในการพัฒนา 38
2.2 โปรแกรม (Software) ที่ใช้ในการพัฒนา 39
2.3 เทคโนโลยีและภาษาคอมพิวเตอร์ที่ใช้ในการพัฒนา 39
3. ข้อมูลที่ใช้ในการศึกษา 39
4. การสืบค้นข้อมูลและการจัดการฐานข้อมูล 41
4.1 การรวบรวมข้อมูล (Data Collection) 41
4.2 การจัดการฐานข้อมูลอุบัติเหตุ (Accident Database) 41
5. การประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ 42
5.1 การประมวลผลภาษาธรรมชาติจากข้อมูลสื่อสังคมออนไลน์ 42
6. การหาความสัมพันธ์ระหว่างข้อมูลจากการประมวลผลภาษาธรรมชาติ 44
7. กระบวนการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์ 44
7.1 การนำเข้าโมดูล เครื่องมือที่สำคัญในการทำ Geocoding 44
7.2 การเตรียมข้อมูลที่ทำการบันทึกจาก Twitter 45
7.3 การเรียกใช้เครื่องมือโมดูล geocode ใน Python 45
7.4 ผลลัพธ์ที่ได้จากการประมวลผล Geocoding 45
8. การจัดทำแผนที่ความเสี่ยงจากการเกิดอุบัติเหตุ ตามช่วงเวลา 46

9. การตรวจสอบความสอดคล้องกับข้อมูลสถิติการเกิดอุบัติเหตุ 46
บทที่ 4 ผลการดำเนินงานวิจัย 47
1. ฐานข้อมูลการเกิดอุบัติเหตุ (Accident Database) 47
2. การประมวลภาษาธรรมชาติและการวิเคราะห์เหมืองดัชนีถ้อยคำของข้อมูลการเกิดอุบัติเหตุ (Natural Language Processing of Accident in social media) 50
2.1 ผลลัพธ์การประมวลผลภาษาธรรมชาติ 50
2.2 การทำความสะอาดข้อมูลด้วยนิพจน์ปรกติ (Regular expression) 51
2.4. การจัดทำกลุ่มข้อมูลและความถี่ข้อมูล 52
3. การวิเคราะห์เหมืองข้อมูลกับสถิติการเกิดอุบัติเหตุจากกระทรวงคมนาคม 62
4. การวิเคราะห์เหมืองข้อมูลและการหาความสัมพันธ์ระหว่างข้อมูลการเกิดอุบัติเหตุ ด้วยอัลกอริทึมอพริโอริ (Association Rule Mining using Apriori Algorithm) 64
5. ตำแหน่งของข้อมูลการเกิดอุบัติเหตุด้วยการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์ (Location of Accident using Geocoding) 70
บทที่ 5 สรุปผล อภิปรายผล และข้อเสนอแนะ 1
1 สรุปผล 1
2 อภิปรายผล 2
2.1 การบันทึกข้อมูลจากสื่อสังคมออนไลน์เพื่อนำมาทำเป็นฐานข้อมูล 2
2.2 การประมวลผลภาษาธรรมชาติและการตัดคำด้วย PythaiNLP 2
2.3 การหาความสัมพันธ์ของข้อความด้วยอัลกอริทึมอพริโอริ 3
3 ข้อเสนอแนะ 3
บรรณานุกรม 5
ประวัติผู้เขียน 10

สารบัญตาราง

หน้า

ตาราง 1 : รายชื่อเขตกรุงเทพมหานคร 7

ตาราง 2 : ถนนที่มีการจราจรหนาแน่นสูงสุด ในช่วงเช้า เวลา 8.30 น. 12

ตาราง 3 : ถนนที่มีการจราจรหนาแน่นสูงสุด ในช่วงกลางวัน เวลา 12.30 น. 12

ตาราง 4 : ถนนที่มีการจราจรหนาแน่นสูงสุด ในช่วงเย็น เวลา 17.30 น. 13

ตาราง 5 ข้อมูลที่ใช้ในการวิจัย 40

ตาราง 6 : คำที่เกี่ยวข้องกับคำสำคัญ (keyword) และค่ากฎความสัมพันธ์ของคำสำคัญ (associate's rule) 64

สารบัญรูปภาพ

หน้า

ภาพประกอบ 1 แผนที่เขต กรุงเทพมหานคร 3
ภาพประกอบ 2 กรอบแนวคิดงานวิจัย 6
ภาพประกอบ 3 แผนที่การใช้ประโยชน์ที่ดิน ปี 2562 10
ภาพประกอบ 4 แผนที่ถนนในกรุงเทพมหานคร 12
ภาพประกอบ 5 ลักษณะการตรวจสอบโครงสร้างทางไวยากรณ์ 18
ภาพประกอบ 6 โครงสร้างการวิเคราะห์ภาษาธรรมชาติ 19
ภาพประกอบ 7 การสกัดข้อความที่ต้องการนำไปวิเคราะห์ 20
ภาพประกอบ 8 ข้อมูลที่อยู่ในรูปแบบถุงข้อความและดัชนีถ้อยคำก้อนเมฆ 21
ภาพประกอบ 9 การจัดกลุ่มของข้อมูล 22
ภาพประกอบ 10 การหาความสัมพันธ์ของข้อมูล 23
ภาพประกอบ 11 อัลกอริทึมอพริโอริ 24
ภาพประกอบ 12 องค์ประกอบของระบบสารสนเทศภูมิศาสตร์ 27
ภาพประกอบ 13 แสดงรายละเอียดข้อมูลสารสนเทศภูมิศาสตร์ 29
ภาพประกอบ 14 แบบจำลองฐานข้อมูล 30
ภาพประกอบ 15 การแปลงข้อมูลที่อยู่ให้เป็นพิกัดทางภูมิศาสตร์ 31
ภาพประกอบ 16 กระบวนการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์ 32
ภาพประกอบ 17 แผนการดำเนินงาน 38
ภาพประกอบ 18 ระบบสื่อสังคมออนไลน์ Twitter บัญชี JS100radio และ fm91trafficpro 40
ภาพประกอบ 19 การเก็บข้อมูลผ่านระบบทวิตเตอร์ 41
ภาพประกอบ 20 ระบบฐานข้อมูลอุบัติเหตุ 41
ภาพประกอบ 21 การเตรียมข้อมูล 42

ภาพประกอบ 22 การกรองข้อมูลด้วย Regular Expression 43
ภาพประกอบ 23 รูปแบบข้อมูลก้อนเมฆ (Word Cloud) 43
ภาพประกอบ 24 การเลือกโมดูลในการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์ 45
ภาพประกอบ 25 การนำข้อมูลมาทำการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์ 45
ภาพประกอบ 26 การเรียกใช้ฟังก์ชันการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์ 45
ภาพประกอบ 27 ผลลัพธ์จากการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์ 45
ภาพประกอบ 28 ระบบฐานข้อมูล 48
ภาพประกอบ 29 การแสดงผลการบันทึกข้อมูลที่ได้จาก Twitter และแสดงผลบนบราวเซอร์ 48
ภาพประกอบ 30 การเชื่อมต่อข้อมูลจาก MySQL เพื่อนำมาประมวลผลภาษาธรรมชาติ 49
ภาพประกอบ 31 ตัวอย่างคลังฐานข้อมูลในรูปแบบ MySQL จากข้อความบน Twitter ในปี 2564-2565 50
ภาพประกอบ 32 การประมวลผลภาษาธรรมชาติ 50
ภาพประกอบ 33 กระบวนการตัดประโยค 51
ภาพประกอบ 34 เปรียบเทียบระหว่างข้อมูลที่ยังไม่ถูกตัดคำ (A) และข้อมูลที่ตัดคำที่ไม่ต้องการด้วย Regular Expression (B) 51
ภาพประกอบ 35 ความถี่ถนนที่มีการเกิดอุบัติเหตุสูงสุดแสดงผลในรูปแบบถ้อยคำก้อนเมฆ 52
ภาพประกอบ 36 กราฟแสดงความถี่ของถนนที่เกิดอุบัติเหตุสูงสุด 53
ภาพประกอบ 37 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ถนนที่เกิดอุบัติเหตุสูงสุด 53
ภาพประกอบ 38 กราฟแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุ 54
ภาพประกอบ 39 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุ 54
ภาพประกอบ 40 กราฟแสดงความถี่ถนนที่เกิดอุบัติเหตุสูงสุดในช่วงเวลา 06:00 – 09:00 น. 55
ภาพประกอบ 41 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ถนนที่เกิดอุบัติเหตุสูงสุดในช่วงเวลา 06:00 – 09:00 น. 56

ภาพประกอบ 42 กราฟแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุในช่วงเวลา 06:00-09:00 น. 56

ภาพประกอบ 43 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุในช่วงเวลา 06:00-09:00 น. 57

ภาพประกอบ 44 กราฟแสดงความถี่ถนนที่เกิดอุบัติเหตุสูงสุดในช่วงเวลา 11:00 – 14:00 น..... 58

ภาพประกอบ 45 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ถนนที่เกิดอุบัติเหตุสูงสุดในช่วงเวลา 11:00 – 14:00 น..... 58

ภาพประกอบ 46 กราฟแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุในช่วงเวลา 11:00-14:00 น. 59

ภาพประกอบ 47 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุในช่วงเวลา 11:00-14:00 น. 59

ภาพประกอบ 48 กราฟแสดงความถี่ถนนที่เกิดอุบัติเหตุสูงสุดในช่วงเวลา 16:00 – 20:00 น..... 60

ภาพประกอบ 49 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ถนนที่เกิดอุบัติเหตุสูงสุดในช่วงเวลา 16:00 – 20:00 น..... 61

ภาพประกอบ 50 กราฟแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุในช่วงเวลา 16:00-20:00 น. 61

ภาพประกอบ 51 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุในช่วงเวลา 16:00-20:00 น. 62

ภาพประกอบ 52 สถิติการเกิดอุบัติเหตุ ปี 2564 ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงคมนาคม 63

ภาพประกอบ 53 ถนนที่มีความถี่ของการเกิดอุบัติเหตุสูงสุดในช่วงเช้า 06.00 ถึง 09.00 71

ภาพประกอบ 54 ถนนที่มีความถี่ของการเกิดอุบัติเหตุสูงสุดในช่วงกลางวัน 11.00 ถึง 14.00 72

ภาพประกอบ 55 ถนนที่มีความถี่ของการเกิดอุบัติเหตุสูงสุดในช่วงเย็น 16:00-20:00..... 73

บทที่1
บทนำ

1. ที่มาและความสำคัญ

ปัจจุบันในกรุงเทพมหานครมีปริมาณการเพิ่มขึ้นของยานพาหนะมากขึ้น ส่งผลให้การจราจรบนท้องถนนคับคั่งเนื่องจากการเพิ่มขึ้นของจำนวนประชากรมีย้ายถิ่นที่อยู่เข้าสู่เมือง ประกอบกับการที่ระบบขนส่งสาธารณะและโครงข่ายถนนมี ไม่เพียงพอและขาดการเชื่อมโยงที่ดี อีกทั้งการขยายตัวของเมืองขาดการควบคุม มีการขยายตัวของเมืองที่เติบโตขึ้นอย่างต่อเนื่อง และการเติบโตของเมืองที่ใหญ่ขึ้นก็จะมากับประชากรที่มีปริมาณมากขึ้นด้วย

การพัฒนาทางเทคโนโลยีสารสนเทศเพื่อใช้ในการติดต่อสื่อสารระหว่างกันเป็นสิ่งสำคัญ เครือข่ายของผู้ให้บริการทางอินเทอร์เน็ต ทำให้การสื่อสารสามารถเข้าถึงได้ทุกคนทุกที่และทุกเวลา ในการติดต่อเชื่อมโยง ผู้ใช้งานทั่วทุกมุมโลกสามารถติดต่อ แลกเปลี่ยนและร่วมแสดงความคิดเห็นต่อข้อมูลเหตุการณ์และกิจกรรมที่สนใจที่มีความหลากหลาย ผ่านรูปแบบตัวอักษร รูปแบบเสียง รูปแบบภาพ หรือรูปแบบวิดิโอ นอกจากนี้ผู้ใช้งานระบบยังสามารถแบ่งปันข้อมูลจากการบริการการบอกตำแหน่งทางภูมิศาสตร์ (Location Based Service: LBS) ซึ่งสามารถใช้ในการตรวจสอบ ติดตาม และแสดงแนวโน้ม เพื่อให้เข้าถึง วิเคราะห์ และประมวลผลข้อมูลที่เกี่ยวข้องด้วยเวลาอันรวดเร็ว (สิทธิ, 2561)

ปัจจุบันการรายงานอุบัติเหตุมีการรายงานผ่านระบบสื่อสังคมออนไลน์ ที่ทำให้ผู้ใช้งานสามารถรับรู้ได้อย่างรวดเร็ว เช่น เฟสบุ๊ค (Facebook) อินสตาแกรม (Instagram) กูเกิ้ลแมพ (Google Maps) หรือ ทวิตเตอร์ (Twitter) เพื่อรับรู้ข้อมูลข่าวสารของสภาพการจราจรและอุบัติเหตุ เพื่อที่จะหลีกเลี่ยงจุดที่ปริมาณการจราจรหนาแน่นหรือเกิดอุบัติเหตุเพื่อลดปริมาณรถยนต์ที่ผ่านจุดเกิดเหตุ ทำให้สภาพการจราจรมีสภาวะที่ดีการนำระบบการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) มาประยุกต์ใช้เพื่อการรู้จำชื่อเฉพาะ (วิกาหะ, 2557)

การพัฒนาระบบการขนส่งและการจราจรอัจฉริยะจะมีลักษณะแตกต่าง กันในแต่ละประเทศ ขึ้นอยู่กับวิธีการพัฒนาและการจัดการขนส่งและการจราจร ซึ่งทำได้หลายแนวทาง หลายรูปแบบ และแตกต่างกันไป สำหรับประเทศไทย นั้น ได้มีแนวคิดที่จะนำระบบการขนส่งและการจราจรอัจฉริยะเข้ามาใช้โดยความ ร่วมมือจากภาครัฐและคณะกรรมการจัดการระบบระบบการขนส่งและการจราจร อัจฉริยะของประเทศไทยได้กำหนดกรอบทิศทางการพัฒนาระบบระบบการขนส่ง และการจราจรอัจฉริยะ เช่น การจัดทำระบบรายงานจราจรแบบตามเวลาจริง (Real-

time),การติดตั้งโทรทัศน์วงจรปิด (CCTV) ให้ครอบคลุมพื้นที่กรุงเทพและ ปริมณฑล การจัดตั้งศูนย์ให้บริการข้อมูลจราจร รวมถึงแผนการติดตั้งระบบถ่าย ภาพผู้ฝ่าฝืนสัญญาณไฟแดงบริเวณทางแยก นอกจากนี้ยังมีระบบอื่น ๆ ที่จะเอื้อ ประโยชน์ต่อการใช้งาน ได้แก่ เครือข่ายเซ็นเซอร์ไร้สายสำหรับตรวจนับรถยนต์, ระบบนับสัญญาณเวลาไฟ, ระบบกระจายข้อมูลสภาพการจราจรแบบ Realtime และโปรแกรมรู้จำป้ายทะเบียนรถ ระบบต่าง ๆ เหล่านี้ล้วนสามารถเชื่อม โยงและพัฒนาให้สามารถทำงานได้โดยซอฟต์แวร์ประมวลผลและรายงานผล ที่ ต้องอาศัยข้อมูลจากแหล่งต่าง ๆ จากหลายหน่วยงานมาใช้เพื่อให้เกิดการพัฒนา ระบบอัจฉริยะสำหรับการเดินทางที่เกิดขึ้นในอนาคตโดยหน่วยงานที่เกี่ยวข้อง สามารถนำข้อมูลที่มีอยู่ทั้งหมดเข้าสู่เครือข่ายต่างๆ และเชื่อมโยงข้อมูลผ่านระบบ อินเทอร์เน็ตเพื่อการประสานงาน และการติดต่อด้วยข้อมูลที่ทันสมัย ซึ่งในหลาย หน่วยงานก็มีแนวทางที่จะนำระบบการขนส่งและการจราจรอัจฉริยะมาพัฒนาให้ สามารถใช้งานได้อย่างเป็นระบบ (คำภีรานนท์, 2558)

ในงานวิจัยนี้ศึกษาการรู้จำสถานที่ในข้อความภาษาไทย จากการรายงานสภาพการจราจรจากสื่อสังคมออนไลน์ในรูปแบบประโยคต่างๆ ประยุกต์การประมวลผลภาษาธรรมชาติเพื่อนำไปเป็นแนวทางแก้ไขปัญหาอุบัติเหตุจราจรอีกแนวทางหนึ่ง ในการประเมินสภาพการจราจรและการระบุตำแหน่งของอุบัติเหตุจากระบบสื่อสังคมออนไลน์ เพื่อการประมาณตำแหน่งพื้นที่เสี่ยงในการเกิดอุบัติเหตุในกรุงเทพมหานคร ทำให้สามารถประมาณการเกิดอุบัติเหตุ สาเหตุของการเกิดอุบัติเหตุ จากการรายงานผ่านระบบสื่อสังคมออนไลน์อย่างรวดเร็ว รวมทั้งสามารถบ่งบอกถึงช่วงเวลา ประเภทอุบัติเหตุ เพื่อเป็นประโยชน์ในการรายงานเหตุการณ์อุบัติเหตุแบบเรียลไทม์ได้ในอนาคต

2. วัตถุประสงค์ของงานวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อคาดการณ์อุบัติเหตุและความเสี่ยงของการเกิดอุบัติเหตุบนท้องถนนโดยมีการแบ่งการศึกษาเป็น 3 วัถตุประสงค์ ได้แก่

1. สร้างฐานข้อมูลการเกิดอุบัติเหตุเชิงพื้นที่จากสื่อสังคมออนไลน์

2. วิเคราะห์ความสัมพันธ์ระหว่างข้อความที่เกี่ยวข้องกับอุบัติเหตุด้วยการประมวลผลภาษาธรรมชาติ

3. การประมาณค่าเชิงตำแหน่งเพื่อวิเคราะห์ความถี่ของอุบัติเหตุด้วยการประมวลผลภาษาธรรมชาติ

3. ความมุ่งหมายของงานวิจัย

1. เพื่อให้ได้ระบบฐานข้อมูลเชิงพื้นที่จากข้อมูลสื่อสังคมออนไลน์ที่เกี่ยวข้องกับการเกิดอุบัติเหตุ

2. เพื่อประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติและการหาความสัมพันธ์ของเหมืองดัชนีถ้อยคำจากสื่อสังคมออนไลน์ที่เกี่ยวข้องกับการเกิดอุบัติเหตุ

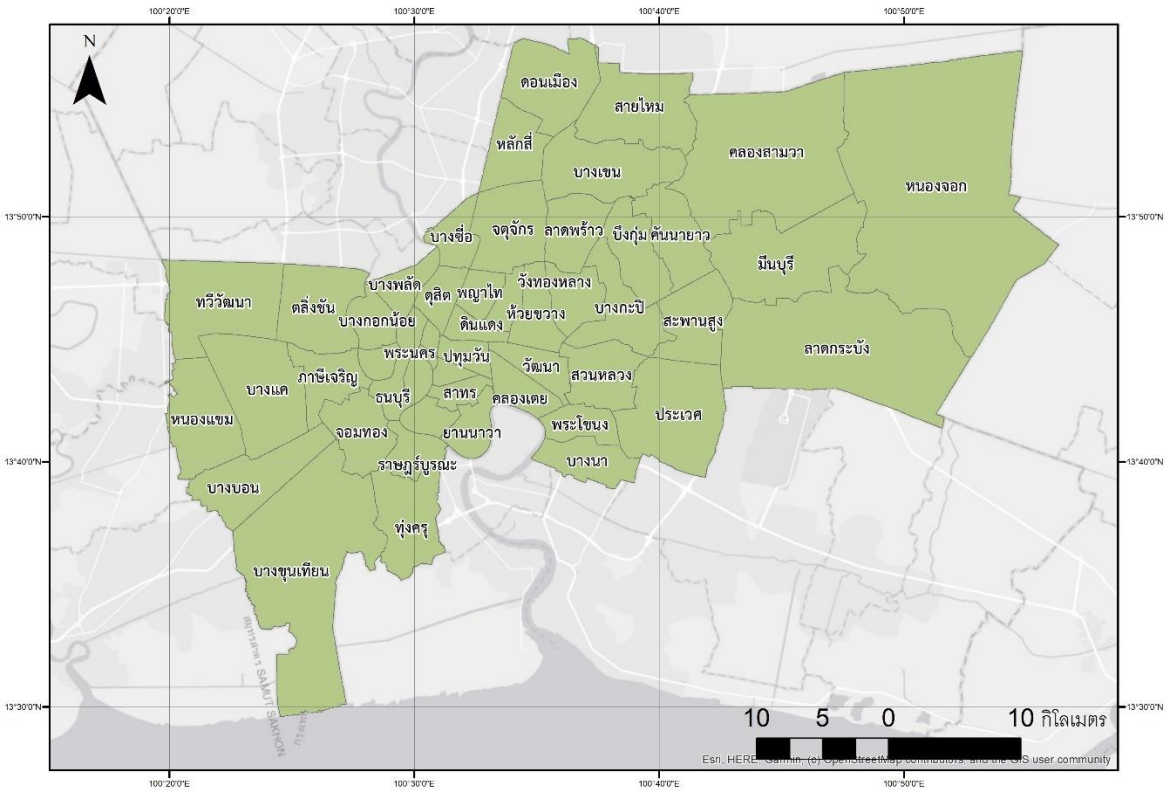
3. เพื่อพัฒนาระบบการประมาณค่าเชิงตำแหน่งการเกิดอุบัติเหตุด้วยการประมวลผลภาษาธรรมชาติเพื่อนำมาแสดงผลเป็นแผนที่การเกิดอุบัติเหตุ

4. ขอบเขตของงานวิจัย

การศึกษาได้กำหนดขอบเขตของการศึกษาไว้ ดังนี้

1. ขอบเขตด้านพื้นที่

กรุงเทพมหานครที่เมืองหลวงและนครที่มีประชากรมากที่สุดของประเทศไทย และเป็นจังหวัดที่มีปริมาณการจราจรมากที่สุดในประเทศไทย จึงเป็นสาเหตุให้มีความเสี่ยงต่อการเกิดอุบัติเหตุบนท้องถนนได้ แบ่งออกเป็น 50 เขต ดังภาพประกอบ 1



ภาพประกอบ 1 แผนที่เขต กรุงเทพมหานคร

2. ขอบเขตด้านเนื้อหา โดยมีการศึกษาดังนี้

การรายงานอุบัติเหตุจากสื่อสังคมออนไลน์ Platform Twitter โดยใช้ข้อมูลจากบัญชีผู้ใช้งาน 2 บัญชี คือ จส.100 (JS100Radio) และ สวพ.91 (fm91trafficpro) เนื่องจากเป็นบัญชีผู้ใช้งานที่รายงานข้อมูล เกี่ยวกับการจราจร ที่มีความน่าเชื่อถือ และเป็นที่นิยม ใช้งานกันอย่างแพร่หลายในประเทสไทย ซึ่งสามารถเป็นคลังข้อมูล เพื่อนำไปวิเคราะห์และเป็นที่กล่าวถึงเมื่อต้องการทราบข้อมูลที่เกี่ยวกับการจราจร เป็นคลังฐานข้อมูลเพื่อนำเหมืองดัชนีข้อความ ไปประมวลผลภาษาธรรมชาติและกระบวนการแปลงข้อความเป็นระบบพิกัดภูมิศาสตร์ เพื่อคาดการณ์การเกิดอุบัติเหตุเชิงพื้นที่

1. ขอบเขตเชิงเวลา มีการแบ่งช่วงเวลาการศึกษา ดังนี้

การรายงานอุบัติเหตุจากสื่อสังคมออนไลน์ ในช่วงเดือนมีนาคม 2564 – มีนาคม2565 โดยแบ่งช่วงเวลาของการรายงานอุบัติเหตุเป็น 3 ช่วงเวลา ได้แก่ ช่วงเช้า เวลา 06:00-09:00 น. ช่วงกลางวัน เวลา 11:00-14:00 น. และช่วงเย็น เวลา 16:00-20:00 น. เนื่องจากเป็นช่วงเวลา ที่มีการจราจรหนาแน่น (TomTom, 2022) จึงเป็นสาเหตุให้ช่วงเวลาดังกล่าวมีโอกาสการเกิดอุบัติเหตุสูงกว่าช่วงเวลาอื่น

2. ขอบเขตประชากร

การรายงานอุบัติเหตุจากสื่อสังคมออนไลน์จะกำหนดขอบเขตของประชากรผู้ใช้งานระบบสื่อสังคมออนไลน์ แอพลิเคชันทวิตเตอร์ จากการรายงานสถิติข้อมูลจากเวปไซต์ datareportal มีจำนวนผู้ใช้งานมากกว่า 14.60 ล้านบัญชี หรือประมาณ 20.3 เปอร์เซ็นต์ของจำนวนประชากรในประเทศไทย

5. **คำสำคัญ**

อุบัติเหตุ สื่อสังคมออนไลน์ การประมวลผลภาษาธรรมชาติ การระบุตำแหน่งด้วยระบบพิกัดทางภูมิศาสตร์

6. **นิยามศัพท์เฉพาะ**

อุบัติเหตุจากการจราจร (Traffic Accident) หมายถึงเหตุการณ์อันตรายที่เกิดขึ้นโดยไม่คาดคิด เหตุการณ์ที่ไม่มีเจตนาให้เกิด เมื่อเกิดขึ้นแล้วมีผลทำให้เกิดการบาดเจ็บพฤติกรรมเสี่ยง ที่มักเกิดขึ้นโดยผู้กระทำไม่ทันคิด หรือบางครั้งอาจจะกระทำโดยไม่ทันคิดหรือรู้ตัว ที่พบเห็นบ่อยก็อย่างเช่น หันหรือก้มไปหยิบจับสิ่งใด การเล่นโทรศัพท์ หรือมีเด็กอยู่ในรถด้วย พฤติกรรมเหล่านี้จะทำให้ความสนใจของผู้กระทำมักถูกดึงไปจากภาพตรงหน้าหรือถนนโดยไม่ตั้งใจ ซึ่งก็เรียกง่ายๆ

ว่าขาดสตินั่นเอง พฤติกรรมเหล่านี้มักนำมาซึ่งอุบัติเหตุได้ง่ายที่สุด สาเหตุของอุบัติเหตุและการขับรถตามกฎจราจรโดยถ่องแท้แล้ว ปรากฏว่าอุบัติเหตุจราจรทางบกไม่ได้เกิดจากเคราะห์กรรมแต่อย่างใด แต่เกิดจากพฤติกรรมหรือการกระทำของคนเป็นส่วนใหญ่ ซึ่งสรุปสาเหตุได้ดังนี้คือ สาเหตุจากบุคคล คนขับรถ เช่น ขับโดยประมาท, ขับรถเร็ว, ขับรถขณะมึนเมา, ใช้โทรศัพท์ขณะขับรถ, คนเดินถนนและข้ามถนน เช่น ไม่ข้ามถนนตรงทางม้าลายหรือสะพานลอย ข้ามตัดหน้ารถระยะกระชั้นชิด สาเหตุจากรถ เช่น การนำรถที่อุปกรณ์บกพร่องมาใช้ในทาง เช่น เบรก, ไฟสัญญาณ, กระจกส่องหลัง, ที่ปัดน้ำฝน สาเหตุจากทางและเครื่องหมายสัญญาณ เช่น บริเวณทางแยก, ทางโค้ง, ทางชำรุด, เครื่องหมายสัญญาณชำรุด, ไม่มี อุปกรณ์จราจร เพื่อช่วยในการขับขี่ สาเหตุจากภัยธรรมชาติ เช่น ฝนตกหนัก, หมอกลงจัด เป็นต้น (งานป้องกันและบรรเทาสาธารณภัย, 2553)

การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) เป็นวิทยาการแขนงหนึ่งในหมวดหมู่ของเทคโนโลยีปัญญาประดิษฐ์หรือ artificial intelligence ซึ่งช่วยให้คอมพิวเตอร์สามารถเข้าใจ ตลอดจนตีความและใช้งานภาษาปกติที่มนุษย์ใช้สื่อสารได้ โดยเทคโนโลยี NLP นี้ มีรากฐานจากวิทยาการหลากหลายสาขาด้วยกัน โดยเฉพาะด้านวิทยาการคอมพิวเตอร์ (computer science) และภาษาศาสตร์เชิงคำนวณ (computational linguistics) เพื่อวัตถุประสงค์ในการปิดช่องว่างทางการสื่อสารระหว่างมนุษย์และระบบคอมพิวเตอร์ (Institute, 2020)

สื่อสังคมออนไลน์ (Social Media) หมายถึงสื่อที่ผู้ส่งสารแบ่งปันสาร ซึ่งอยู่ในรูปแบบต่างๆ ไปยังผู้รับสารผ่านเครือข่ายออนไลน์โดยสามารถโต้ตอบกันระหว่างผู้ส่งสารและผู้รับสาร หรือผู้รับสารด้วยกันเอง ซึ่งสามารถแบ่งสื่อสังคมออนไลน์ออกเป็นประเภทต่างๆ ที่ใช้กันบ่อยๆ คือบล็อก (Blogging) ทวิตเตอร์และไมโครบล็อก (Twitter and Microblogging) เครือข่ายสังคมออนไลน์ (Social Networking) และการแบ่งปันสื่อทางออนไลน์ (Media Sharing) (วิจิตรบุญยรักษ์, 2553)

กระบวนการแปลงข้อความเป็นพิกัดทางภูมิศาสตร์ (Geocoding) คือ กระบวนการแปลงข้อมูลข้อความไม่ว่าจะเป็น ชื่อ หรือที่อยู่ (Address) ให้เป็นตำแหน่ง พิกัดภูมิศาสตร์ อ้างอิง บนพื้นผิวโลก เช่น เส้นรุ้ง (Latitude) เส้นแวง (Longitude) ที่เราสามารถใช้ในการระบุตำแหน่งเพื่อกำหนดจุดลงบนแผนที่ ลดจำนวนข้อมูลตัวอักษรในการระบุที่อยู่ บ้านเลขที่ ถนน หมู่บ้าน ตำบล (Surapong, 2019)

7. ประโยชน์ที่คาดว่าจะได้รับ

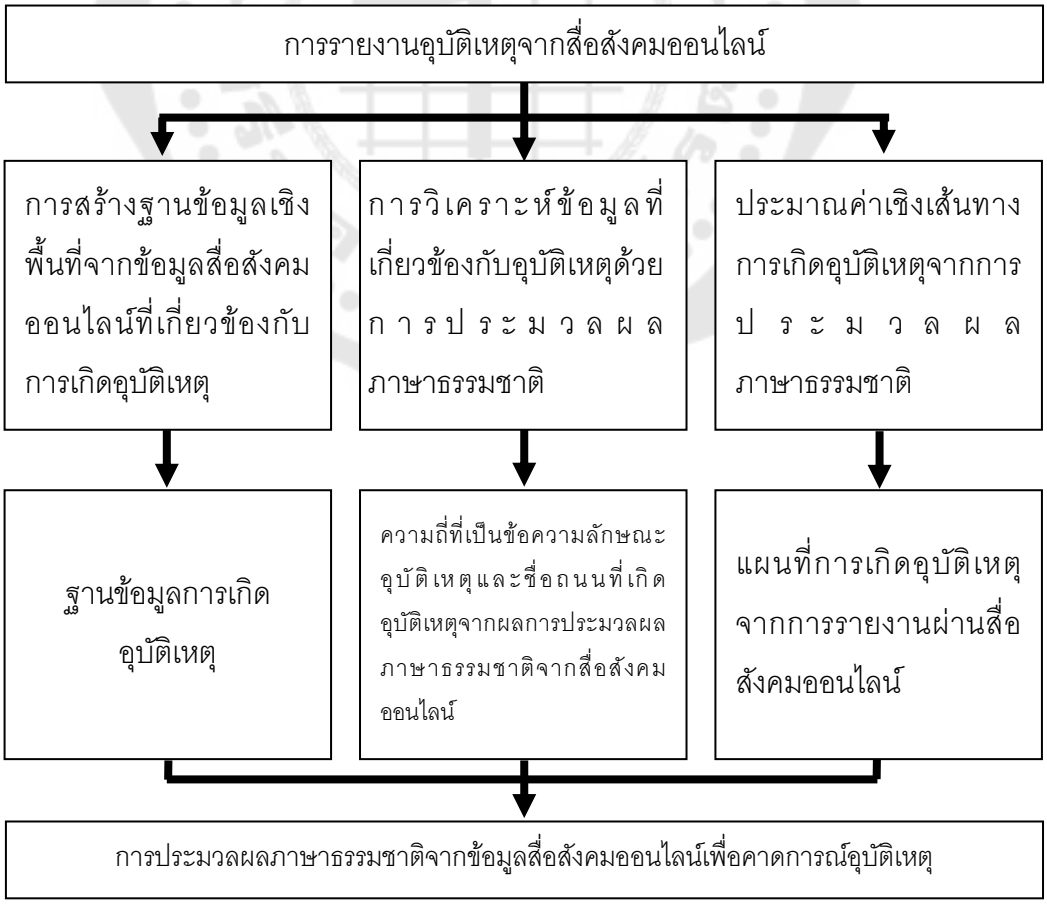
1. เพื่อให้ระบบการรวบรวมและจัดทำฐานข้อมูลจากกระบวนการประมวลผลภาษาธรรมชาติมีโครงสร้างที่แน่นอนและสามารถนำไปวิเคราะห์ลักษณะการเกิดอุบัติเหตุเชิงพื้นที่ได้

2. เพื่อทำให้การประมวลผลภาษาธรรมชาติ ด้วยภาษาไทยมีความแม่นยำ โดยการนำข้อมูลอุบัติเหตุจากสื่อสังคมออนไลน์มาทำการวิเคราะห์

3. เพื่อทำให้การแปลงระบบพิกัดทางภูมิศาสตร์ของการเกิดอุบัติเหตุจากการประมวลผลภาษาธรรมชาติมีความถูกต้องและแม่นยำมากขึ้น

8. กรอบแนวคิดงานวิจัย

งานวิจัยนี้แบ่งลักษณะวิธีออกเป็น 5 กระบวนการหลัก ได้แก่ การรวบรวมข้อมูล การจัดการฐานข้อมูล การประมวลผลข้อความ การหาความสัมพันธ์ระหว่างข้อมูล การจัดทำแผนที่ความเสี่ยงจากการเกิดอุบัติเหตุโดยมีกรอบแนวคิดของงานวิจัย ดังภาพประกอบ 2



ภาพประกอบ 2 กรอบแนวคิดงานวิจัย

บทที่2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การประมวลผลภาษาธรรมชาติจากข้อมูลสื่อสังคมออนไลน์เชิงพื้นที่เพื่อคาดการณ์อุบัติเหตุ ได้ศึกษาข้อมูลเอกสารและงานวิจัยที่เกี่ยวข้อง มีแนวคิด ทฤษฎีเอกสาร และงานวิจัยที่เกี่ยวข้อง มีหัวข้อในการศึกษาดังนี้

1. ข้อมูลทั่วไปของพื้นที่ศึกษา
2. ระบบสื่อสังคมออนไลน์ (Social Media)
3. การประมวลผลภาษาธรรมชาติ (Natural Languages Processing : NLP)
4. ระบบสารสนเทศภูมิศาสตร์ (Geographic Information System : GIS)
5. เอกสารและงานวิจัยที่เกี่ยวข้อง

1 ข้อมูลทั่วไปของพื้นที่ศึกษา

กรมพัฒนาที่ดิน (2562) ได้อธิบายว่า กรุงเทพมหานคร มีตำแหน่งทางภูมิศาสตร์ในเขตภาคกลางของประเทศไทย โดยการแบ่งภูมิภาคประเทศไทยได้ใช้เกณฑ์ด้านภูมิศาสตร์ของราชบัณฑิตยสถาน ซึ่งจากการคำนวณเนื้อที่ด้วยระบบสารสนเทศภูมิศาสตร์ พบว่า กรุงเทพมหานคร มีเนื้อที่ 1,569 ตารากิโลเมตร หรือ 981,631 ไร่ เป็นเมืองหลวงและนครที่มีประชากรมากที่สุดของประเทศไทย เป็นศูนย์กลางการปกครอง การศึกษา การคมนาคมขนส่ง การเงินการธนาคาร การพาณิชย์ การสื่อสาร และความเจริญของประเทศ ตั้งอยู่บนสามเหลี่ยมปากแม่น้ำเจ้าพระยา มีแม่น้ำเจ้าพระยาไหลผ่านและแบ่งเมืองออกเป็น 2 ฝั่ง คือ ฝั่งพระนครและฝั่งธนบุรี มีประชากรตามทะเบียนราษฎรกว่า 6 ล้านคน ทำให้กรุงเทพมหานครจัดเป็นเอกนคร (Primate City) มีผู้กล่าวว่า กรุงเทพมหานครเป็น "เอกนครที่สุดในโลก" เพราะมีประชากรมากกว่านครที่มีประชากรมากเป็นอันดับ 2 ถึง 40 เท่า ปัจจุบันกรุงเทพมหานครมีเขตการปกครอง 50 เขต ดังตารางที่ 1

ตาราง 1 : รายชื่อเขตกรุงเทพมหานคร

ลำดับ	อำเภอ/กิ่งอำเภอ	เนื้อที่ (ตร.กม.)	ลำดับ	อำเภอ/กิ่งอำเภอ	เนื้อที่ (ตร.กม.)
1	เขตคลองเตย	12.99	26	เขตบึงกุ่ม	24.31
2	เขตคลองสาน	6.05	27	เขตปทุมวัน	8.37

ตาราง 1 (ต่อ)

3	เขตคลองสามวา	110.69	28	เขตประเวศ	52.49
4	เขตคันนายาว	25.98	29	เขตป้อมปราบศัตรูพ่าย	1.93
5	เขตจตุจักร	32.91	30	เขตพญาไท	9.6
6	เขตจอมทอง	26.27	31	เขตพระโขนง	13.99
7	เขตดอนเมือง	36.8	32	เขตพระนคร	5.54
8	เขตดินแดง	8.35	33	เขตภาษีเจริญ	17.83
9	เขตดุสิต	10.67	34	เขตมีนบุรี	63.65
10	เขตตลิ่งชัน	29.48	35	เขตยานนาวา	16.66
11	เขตทวีวัฒนา	50.22	36	เขตราชเทวี	7.13
12	เขตทุ่งครุ	30.74	37	เขตราษฎร์บูรณะ	15.78
13	เขตธนบุรี	8.55	38	เขตลาดกระบัง	123.86
14	เขตบางเขน	42.12	39	เขตลาดพร้าว	21.86
15	เขตบางแค	44.46	40	เขตวังทองหลาง	19.57
16	เขตบางกอกใหญ่	6.18	41	เขตวัฒนา	12.57
17	เขตบางกอกน้อย	11.94	42	เขตสวนหลวง	23.68
18	เขตบางกะปิ	28.52	43	เขตสะพานสูง	28.12
19	เขตบางขุนเทียน	120.69	44	เขตสัมพันธวงศ์	1.42
20	เขตบางคอแหลม	10.92	45	เขตสาทร	9.33
21	เขตบางซื่อ	11.55	46	เขตสายไหม	44.62
22	เขตบางนา	18.79	47	เขตหนองแขม	35.83
23	เขตบางบอน	34.75	48	เขตหนองจอก	236.26
24	เขตบางพลัด	11.36	49	เขตหลักสี่	22.84
25	เขตบางรัก	5.54	50	เขตห้วยขวาง	15.03

ที่มา : http://www.e-report.energy.go.th/area/Bangkok.html

1.1 สภาพภูมิประเทศ

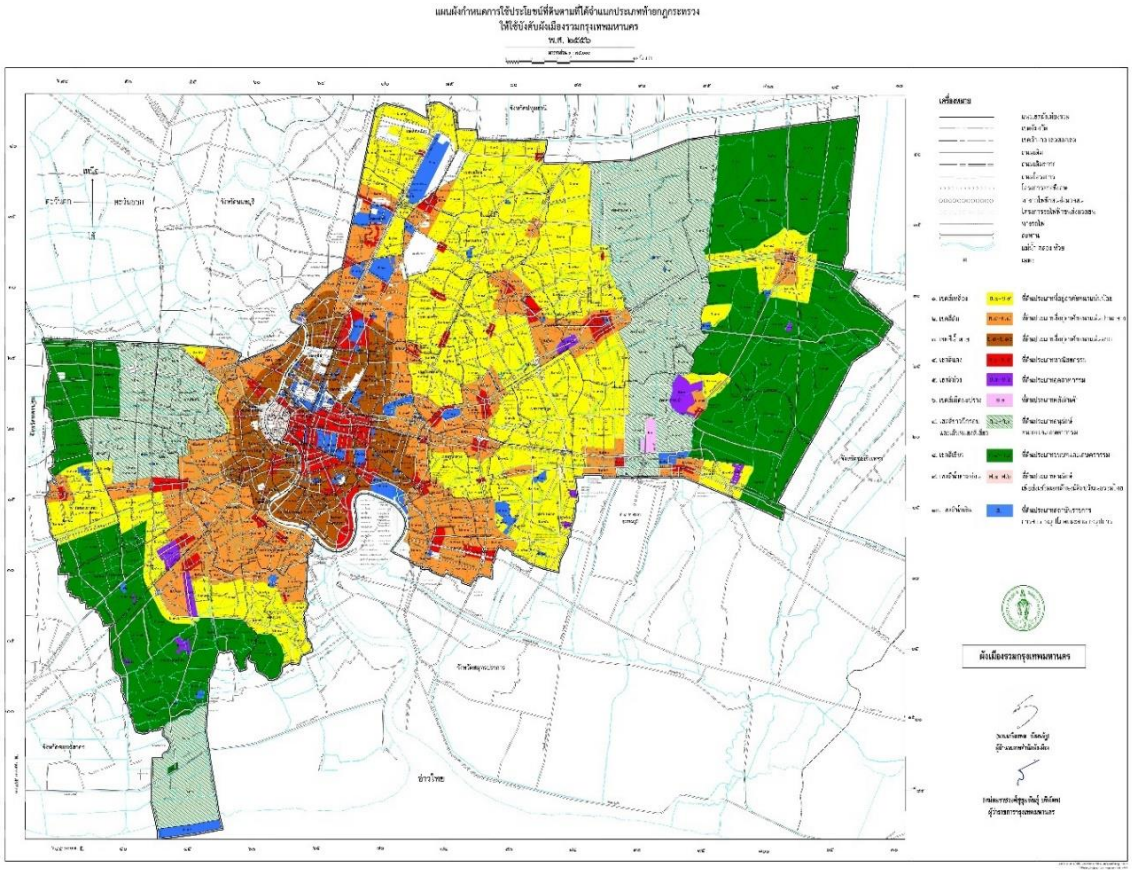
กรุงเทพมหานครมีลักษณะเป็นที่ราบลุ่มหรือที่เรียกว่า “ที่ราบดอนสามเหลี่ยมปากแม่น้ำ” ซึ่งเกิดจากตะกอนน้ำพัดพาสะสม มีระดับความสูงจากระดับน้ำทะเลปานกลางประมาณ 1.50 – 2.00 เมตร มีความลาดเอียงจากทิศเหนือ ค่อยๆลาดเอียงสู่อ่าวไทยทิศใต้ และเฉพาะลุ่มแม้น้ำเจ้าพระยาตอนล่างจะอยู่สูงกว่าระดับน้ำทะเลไม่เกิน 1.50 เมตร โดยเป็นส่วนหนึ่งของที่ราบลุ่มภาคกลางตอนล่างของประเทศไทย (The Lower General Plain of Thailand) เป็นพื้นที่อุดมสมบูรณ์เหมาะแก่การเพาะปลูกข้าวและพืชประเภทต่างๆ

1.2 สภาพภูมิอากาศ

สภาพภูมิอากาศขอกรุงเทพมหานคร อยู่ภายใต้อิทธิพลของลมมรสุม 2 ชนิด คือ 1.ลมมรสุมตะวันออกเฉียงเหนือที่พัดเข้ามาปกคลุมในช่วงฤดูหนาว ทำให้อากาศหนาวเย็นและแห้แล้ง 2.ลมมรสุมตะวันออกเฉียงใต้ที่พัดเข้ามาปกคลุมในช่วงฤดูฝน ทำให้อากาศชุ่มชื้นและมีฝนทั่วไป

1.3 การใช้ประโยชน์ที่ดินจังหวัดกรุงเทพมหานคร

กรุงเทพมหานคร มีเนื้อที่ทั้งหมด 978,253 ไร่ สามารถจำแนกประเภทการใช้ที่ดินได้ออกเป็น 5 ประเภท คือ พื้นที่ชุมชนและสิ่งปลกสร้าง (U) มีเนื้อที่ 659,114 ไร่ คิดเป็นร้อยละ 67.36 เปอร์เซ็นต์ พื้นที่เกษตรกรรม (A) มีเนื้อที่ 231,441 ไร่ คิดเป็นร้อยละ 23.67 เปอร์เซ็นต์ พื้นที่ป่าไม้ (F) มีเนื้อที่ 1,582 ไร่ คิดเป็นร้อยละ 0.16 เปอร์เซ็นต์ พื้นที่น้ำ (W) มีเนื้อที่ 28,821 ไร่ คิดเป็นร้อยละ 2.96 เปอร์เซ็นต์ และพื้นที่เบ็ดเตล็ด (M) มีเนื้อที่ 57,305 ไร่ คิดเป็นร้อยละ 5.85 เปอร์เซ็นต์ ดังภาพประกอบ 3



ภาพประกอบ 3 แผนที่การใช้ประโยชน์ที่ดิน ปี 2562

ที่มา: กรมพัฒนาที่ดิน,2562

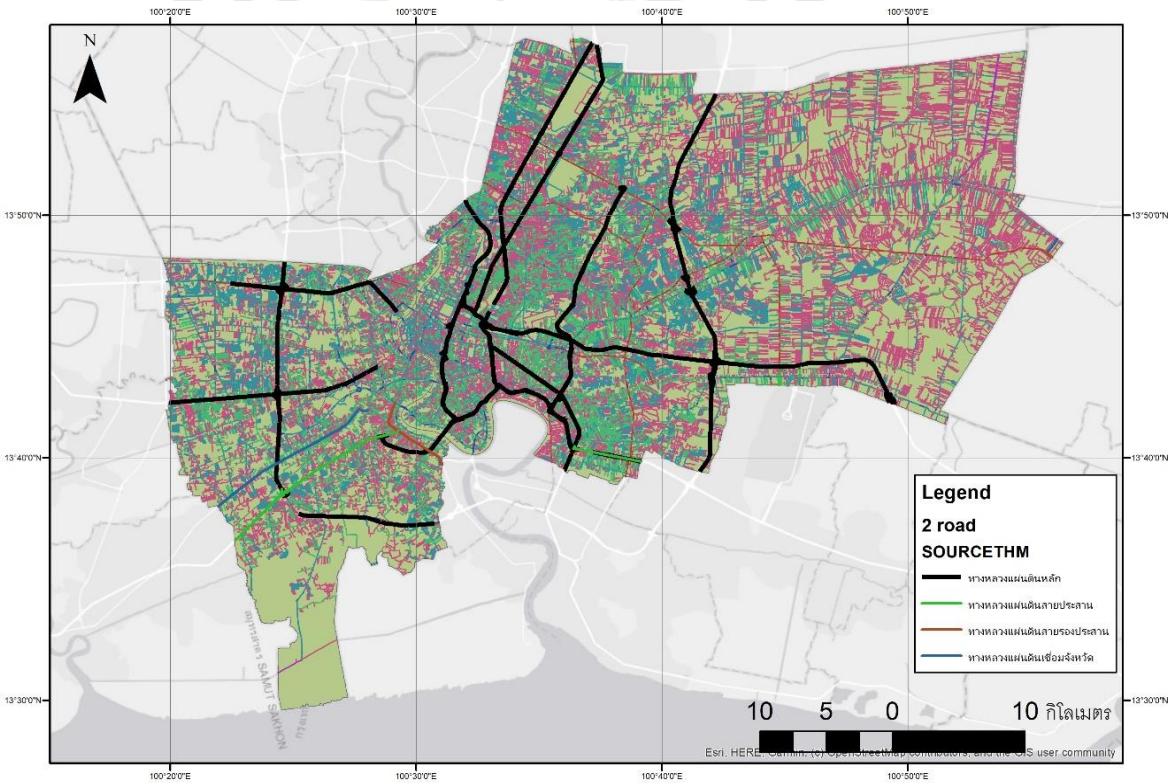
1.4 ระบบคมนาคมขนส่งทางบกในกรุงเทพมหานคร

จากการมีกิจกรรมกระจุกตัวอยู่อย่างหนาแน่นและแรงงานที่อพยพเข้ามาสู่กรุงเทพมหานคร ทำให้เกิดปัญหาระบบโครงข่ายคมนาคมขนส่งไม่สามารถรองรับการเจริญเติบโตและการเพิ่มจำนวนประชากรอย่างรวดเร็ว และตอบสนองอย่างเพียงพอต่อความต้องการของประชาชนที่เดินทางถึง 18 ล้านเที่ยวต่อวัน การเปลี่ยนแปลงการใช้ประโยชน์ที่ดินจากแปลงเกษตรกรรมขนาดใหญ่มาเป็นการใช้ประโยชน์ที่ดินแบบเมือง โดยไม่มีการวางแผนรองรับอย่างเหมาะสม ส่งผลให้พื้นที่ถนนในกรุงเทพฯ มีปริมาณต่ำกว่ามาตรฐานของความเป็นเมืองที่มีการเดินทางอย่างสะดวก กรุงเทพมหานครมีพื้นที่ถนนประมาณ ร้อยละ 10 ของพื้นที่ทั้งเมืองทั้งหมด อีกทั้งประชาชนส่วนมากยังมีความต้องการในการใช้รถยนต์ส่วนตัว ระบบขนส่งมวลชนส่วนใหญ่ยังใช้เส้นทางร่วมกับรถยนต์ส่วนบุคคล ส่วนการเดินทางด้วยระบบขนส่งมวลชนบนรางมีเพียงร้อยละ 3 ของปริมาณการเดินทางทั้งหมด เนื่องจากโครงข่ายยังไม่ครบสมบูรณ์และขาดการวางแผนการจัดการแบบบูรณาการกับการเดินทางประเภทอื่นๆ

ระบบคมนาคมขนส่งของกรุงเทพมหานครในปัจจุบัน เป็นระบบที่ใช้ทางบกเป็นหลัก โดยเฉพาะการคมนาคมขนส่งตามเส้นทางถนน เนื่องจากเข้าถึงชุมชนได้สะดวกกว่าระบบอื่น และเป็นการเดินทางโดยพึ่งยานพาหนะส่วนบุคคล ซึ่งอัตราส่วนการใช้รถยนต์ส่วนบุคคลมีถึงร้อยละ 53 ในขณะที่สัดส่วนการใช้รถขนส่งมวลชนมีเพียงร้อยละ 47 โครงการข่ายถนนและทางด่วน โครงข่ายถนนและทางด่วนในกรุงเทพมหานคร เป็นโครงการข่ายหลักที่สำคัญของการคมนาคมขนส่ง โดยในพื้นที่กรุงเทพมหานครและปริมณฑลมีถนนสายสำคัญประมาณ 4,700 กิโลเมตร ดังภาพประกอบ 4 กรุงเทพมหานครมีถนนสายหลักที่เป็นเส้นทางเข้า-ออกเมืองในทิศต่างๆ ดังนี้

- ด้านทิศเหนือ ได้แก่ ถนนวิภาวดีรังสิต ถนนพหลโยธิน ถนนประชาชื่น ถนนงามวงศ์วาน ถนนรามอินทรา
- ด้านทิศตะวันออก ได้แก่ ถนนสุขุมวิท ถนนเพชรบุรี ถนนพระราม9 ถนนพระราม4 ถนนศรีนครินทร์ ถนนอ่อนนุช
- ด้านทิศตะวันตก ได้แก่ ถนนเพชรเกษม ถนนบรมราชชนนี ถนนจรัญสนิทวงศ์ ถนนพุทธมณฑล ถนนกาญจนาภิเษก(วงแหวนรอบนอก)
- ด้านทิศใต้ ได้แก่ ถนนพระราม2 ถนนพระราม3 ถนนตากสิน ถนนเอกชัย

ที่มา : (สำนักปลัดกรุงเทพมหานคร, 2550)



ภาพประกอบ 4 แผนที่ถนนในกรุงเทพมหานคร

1.5 ถนนที่มีการจราจรหนาแน่น ในจังหวัดกรุงเทพมหานคร

จากสถิติการจราจรโดยใช้ข้อมูลจาก มูลนิธิศูนย์ข้อมูลจราจรอัจฉริยะไทย และบริการ Longdo Traffic เว็บไซต์รายงานสภาพจราจรโดยทำการแบ่งช่วงเวลาของปริมาณการจราจร คือ ช่วงเช้า เวลา 08:30 น. ได้ทำการเก็บสถิติถนนที่มีการจราจรติดขัดสูงสุด ได้แก่ ทางพิเศษศรีรัช ทางพิเศษเฉลิมมหานคร ถนนรัชดาภิเษก ถนนพหลโยธิน ถนนสุขุมวิท ตามลำดับ ดังตาราง 2

ตาราง 2 : ถนนที่มีการจราจรหนาแน่นสูงสุด ในช่วงเช้า เวลา 8.30 น.

ลำดับ	ชื่อถนน	พื้นที่จังหวัด/เขต
1	ทางพิเศษศรีรัช	กรุงเทพมหานคร
2	ทางพิเศษเฉลิมมหานคร	กรุงเทพมหานคร
3	ถนนรัชดาภิเษก	กรุงเทพมหานคร
4	ถนนพหลโยธิน	กรุงเทพมหานคร
5	ถนนสุขุมวิท	กรุงเทพมหานคร
6	ถนนเลียบมอเตอร์เวย์ กรุงเทพ - ชลบุรี	กรุงเทพมหานคร
7	ถนนลาดพร้าว	กรุงเทพมหานคร
8	ทางพิเศษฉลองรัช	กรุงเทพมหานคร
9	ถนนรามอินทรา	กรุงเทพมหานคร
10	ถนนพระราม 4	กรุงเทพมหานคร

ที่มา : https://traffic.longdo.com/bkk-speed ,2565

ช่วงกลางวัน เวลา 12:30 น. ได้ทำการเก็บสถิติถนนที่มีการจราจรติดขัดสูงสุด ได้แก่ ถนนสุขุมวิท ทางพิเศษศรีรัช ถนนลาดพร้าว ถนนรัชดาภิเษก ทางพิเศษเฉลิมมหานคร ตามลำดับ ดังตาราง 3

ตาราง 3 : ถนนที่มีการจราจรหนาแน่นสูงสุด ในช่วงกลางวัน เวลา 12.30 น.

ลำดับ	ชื่อถนน	พื้นที่จังหวัด/เขต
1	ถนนสุขุมวิท	กรุงเทพมหานคร

ตาราง 3 (ต่อ)

2	ทางพิเศษศรีรัช	กรุงเทพมหานคร
3	ถนนลาดพร้าว	กรุงเทพมหานคร
4	ถนนรัชดาภิเษก	กรุงเทพมหานคร
5	ทางพิเศษเฉลิมมหานคร	กรุงเทพมหานคร
6	ถนนเลียบมอเตอร์เวย์	กรุงเทพมหานคร
7	ถนนมอเตอร์เวย์ สาย 7	กรุงเทพมหานคร
8	ถนนกาญจนาภิเษก	กรุงเทพมหานคร
9	ถนนคู้คลองสิบ	กรุงเทพมหานคร
10	ถนนพหลโยธิน	กรุงเทพมหานคร

ที่มา : https://traffic.longdo.com/bkk-speed , 2565

ช่วงเย็น เวลา 17:30 น. ได้ทำการเก็บสถิติถนนที่มีการจราจรติดขัดสูงสุด ได้แก่ ทางพิเศษเฉลิมมหานคร ทางพิเศษศรีรัช ถนนสุขุมวิท ถนนลาดพร้าว ถนนศรีนครินทร์ ตามลำดับ ดังตาราง 4

ตาราง 4 : ถนนที่มีการจราจรหนาแน่นสูงสุด ในช่วงเย็น เวลา 17.30 น.

ลำดับ	ชื่อถนน	พื้นที่จังหวัด/เขต
1	ทางพิเศษเฉลิมมหานคร	กรุงเทพมหานคร
2	ทางพิเศษศรีรัช	กรุงเทพมหานคร
3	ถนนสุขุมวิท	กรุงเทพมหานคร
4	ถนนลาดพร้าว	กรุงเทพมหานคร
5	ถนนศรีนครินทร์	กรุงเทพมหานคร
6	ถนนรัชดาภิเษก	กรุงเทพมหานคร
7	ถนนเลียบมอเตอร์เวย์	กรุงเทพมหานคร
8	ถนนแจ้งวัฒนะ	กรุงเทพมหานคร
9	ทางพิเศษฉลองรัช	กรุงเทพมหานคร

ตาราง 4 (ต่อ)

10	ถนนพระราม 4	กรุงเทพมหานคร

ที่มา : https://traffic.longdo.com/bkk-speed ,2565

กรุงเทพมหานครมีพื้นที่ทั้งหมด 1,569 ตารางกิโลเมตร แบ่งเป็นพื้นที่ถนน 113.06 ตารางกิโลเมตร หรือคิดเป็นเพียงร้อยละ 7.2 ของพื้นที่กรุงเทพมหานครเท่านั้น ซึ่งต่ำกว่าค่ามาตรฐานกว่า 3 เท่า โดยเขตที่มีสัดส่วนพื้นที่ถนนมากที่สุด 5 อันดับคือ เขตป้อมปราบศัตรูพ่าย 16.7% เขตบางรัก 15.4%เขตราชเทวี 14.3% เขตปทุมวัน 14.1% และเขตพระนคร 13.5%

การศึกษาอุบัติเหตุจากการจราจรในกรุงเทพมหานครต้องมีการใช้ข้อมูลเส้นทางโดยในแผนที่แสดงข้อมูลถนน แสดงข้อมูลเฉพาะถนนสาธารณะที่อยู่ในความดูแลของหน่วยงานในสังกัดกรุงเทพมหานคร กรมทางหลวง และกรมทางหลวงชนบท รวมทั้งเส้นทางที่ตัดผ่านที่ดินของหน่วยงานราชการอื่นและเส้นทางของเอกชนที่ได้รับการกำหนดให้เป็นถนนสาธารณะอย่างเป็นทางการจากกรุงเทพมหานครแล้ว ทั้งนี้ ไม่รวมเส้นทางที่ใช้สัญจรเฉพาะภายในพื้นที่หน่วยงานราชการ สถาบันการศึกษา บ้านจัดสรร และสถานที่เอกชนซึ่งมักมีการปิดกั้นหรือควบคุมทางเข้าออก (สำนักงานจราจรและการขนส่ง, 2562)

2. สื่อสังคมออนไลน์ (Social Media)

สื่อสังคมออนไลน์ หมายถึง สื่อดิจิทัลที่เป็นเครื่องมือในการปฏิบัติการทา [คำอธิบายสัญลักษณ์] l Tools) เพื่อใช้สื่อสารระหว่างกันในเครือข่ายทางสังคม (Social Network) ผ่านทางเว็บไซต์และโปรแกรมประยุกต์บนสื่อใดๆ ที่มีการเชื่อมต่อกับอินเทอร์เน็ต โดยเน้นให้ผู้ใช้ทั้งที่เป็นผู้ส่งสารและผู้รับสารมีส่วนร่วม (Collaborative) อย่างสร้างสรรค์ ในการผลิตเนื้อหาขึ้นเอง (User-Generate Content: UGC) ในรูปของข้อมูล ภาพ และเสียง (อรวรรณ วงศ์แก้วโพธิ์ทอง, 2553)

การเลือกทำเลที่ตั้ง ระบบ Micro Blogging และ Micro Sharing หรือที่เรียกกันว่า “บล็อกจิ๋ว” ซึ่งเป็นเว็บเซอร์วิสหรือเว็บไซต์ที่ให้บริการแก่บุคคลทั่วไป สำหรับให้ผู้ใช้บริการเขียนข้อความสั้นๆ ประมาณ 140 ตัวอักษร ที่เรียกว่า “Status” หรือ “Notice” เพื่อแสดงสถานะของตัวเองว่ากำลังทำอะไรอยู่ หรือแจ้งข่าวสารต่างๆ แก่กลุ่มเพื่อนในสังคมออนไลน์ (Online Social Network) ทั้งนี้การกำหนดให้ใช้ข้อมูลในรูปข้อความสั้นๆ ก็เพื่อให้ผู้ใช้ที่เป็นทั้งผู้เขียนและผู้อ่านเข้าใจง่าย ที่นิยมใช้กันอย่างแพร่หลายคือ Twitter (Online Social Network Wikipedia,2010)

เฟสบุ๊ค (Facebook) เป็นบริการเครือข่ายสังคมสัญชาติอเมริกัน สำนักงานใหญ่อยู่ที่เมนโลพาร์ก รัฐแคลิฟอร์เนีย สหรัฐ เฟซบุ๊กก่อตั้งเมื่อวันที่ 4 กุมภาพันธ์ ค.ศ. 2004 โดยมาร์ก

ซักเคอร์เบิร์ก และเพื่อนร่วมห้องในมหาวิทยาลัยฮาร์วาร์ด สมาชิกเพื่อนผู้ก่อตั้ง คือ เอดูอาโด ซาเวอริน, แอนดรูว์ แม็กคอลลัม, ดัสติน มอสโควิส และคริส ฮิวจ์ส ในตอนแรกจำกัดผู้เข้าชมเฉพาะนักศึกษามหาวิทยาลัยฮาร์วาร์ด ภายหลังได้เพิ่มจำนวนมหาวิทยาลัยในพื้นที่บอสตัน ไอวีลีก และมหาวิทยาลัยสแตนฟอร์ด และค่อย ๆ รับรองมหาวิทยาลัยอื่น ๆ ต่อมารับรองโรงเรียนมัธยมศึกษาโดยเฟซบุ๊กอนุญาตให้เยาวชนอายุต่ำกว่า 13 ปี ทั่วโลกสามารถสมัครสมาชิกได้โดยไม่ต้องอ้างอิงหลักฐานใด ๆ ใน ค.ศ. 2020 เฟซบุ๊กอ้างว่ามีผู้ใช้ที่ยังคงใช้งานรายเดือนที่ 2.9 พันล้านคน โดยมีผู้ใช้งานทั่วโลกมากเป็นอันดับ 7 และเป็นโปรแกรมประยุกต์ที่มียอดดาวน์โหลดมากที่สุดในคริสต์ทศวรรษ 2010 เฟซบุ๊กอนุญาตให้ใครก็ได้เข้าสมัครลงทะเบียนกับเฟซบุ๊ก และผู้เป็นสามาชิกของเฟซบุ๊ก นั้นสามารถสร้างพื้นที่ส่วนตัว สำหรับแนะนำตัวเอง ติดต่อสื่อกับเพื่อน ทั้งแบบ ข้อความ ภาพ เสียง และ วีดีโอ โดยผู้ใช้สามารถเลือกที่จะเป็นหรือไม่เป็นเพื่อนกับใครก็ได้ในเฟซบุ๊กผู้ใช้ยังสามารถใช้เฟซบุ๊ก เพื่อร่วมทำกิจกรรมกับผู้ใช้งานท่านอื่นได้เช่น การเขียนข้อความ เล่าเรื่องความรู้สึก แสดงความคิดเห็นเรื่องที่สนใจ โพสต์รูปภาพ โพสต์คลิปวิดีโอ แชทพูดคุย เล่นเกมที่สามารถชวนผู้ใช้งานท่านอื่นมาเล่นกับเราได้ รวมไปถึงทำกิจกรรมอื่นๆ ผ่านแอพลิเคชั่นมีอยู่อย่างมากมาย ซึ่งแอพลิเคชั่นดังกล่าวได้ถูกพัฒนาเข้ามาเพิ่มเติมอยู่เรื่อยๆ แอพลิเคชั่นยังแบ่งออกเป็นหลายหมวดหมู่ เช่น เพื่อความบันเทิง เกม หรือไม่ว่าจะเป็นเชิงธุรกิจ แอพลิเคชั่นของ Facebook ก็มีให้ใช้งานเช่นเดียวกัน ด้วยเหตุนี้ Facebook จึงได้รับความนิยมมากที่สุดในโลก (Inc., 2020)

ทวิตเตอร์ (Twitter) เป็นบริการเครือข่ายสังคมออนไลน์ประเภทไมโครบล็อก (micro blogging) และข้อความที่ส่งถึงกันมีศัพท์เรียกว่า "Tweets" ข้อความที่จะส่งนั้นต้องเป็น plain text เท่านั้นจะแทรกคำสั่งโปรแกรมอะไรไม่ได้ ยกเว้นแต่ hyperlink มายังเว็บเพจของเรา ที่สามารถใส่ไปได้ โดยระบบจะจัดการต่อให้เอง ข้อดีของ Twitter คือ สามารถอัพเดตข้อมูลได้ตามเวลาจริง สามารถใช้ได้กับสื่อหลายชนิด โดยเฉพาะมือถือ สามารถใช้ในการอัพโหลดข้อความหรือตดตามอัพเดตได้ในทุกสถานที่ ทุกเวลา และสามารถใช้เก็บ link เพื่อเชื่อมต่อไปยังเว็บไซต์ที่เก็บภาพหรือวิดีโอได้ แต่ก็มีข้อเสียคือ ไม่สามารถอัพโหลดภาพหรือวิดีโอได้ ทวิตเตอร์ก่อตั้งเมื่อเดือนมีนาคม พ.ศ. 2549 และต่อมาในเดือนกรกฎาคม ก็ได้เปิดตัวเว็บไซต์สังคมออนไลน์ ซึ่งต่อมาได้รับความนิยมจากทั่วโลก โดยมีผู้ใช้ที่ลงทะเบียนเข้าใช้มากกว่า 500 ล้านคนในปี พ.ศ. 2555 รวมไปถึงมีทวีตมากกว่า 340 ล้านทวีตต่อวัน และมีการค้นหาข้อมูลผ่านเว็บไซต์มากกว่า 1,600 ล้านครั้งต่อวัน นับตั้งแต่วันเปิดตัว ทวิตเตอร์เป็นเว็บไซต์ 1 ใน 10 อันดับที่มีผู้เข้าใช้งานมากที่สุด โดยได้รับการขนานนามว่าเป็นเว็บไซต์สำหรับการส่งบริการข้อความสั้น บนอินเทอร์เน็ต ซึ่งในภายหลัง

นอกจากที่จะสามารถทวีตบนเว็บไซต์แล้ว ได้มีการเปิดให้ใช้งานการส่งทวีต ด้วยการส่งบริการข้อความสั้น (SMS) และบนโปรแกรมประยุกต์ ในโทรศัพท์มือถือ และสมาร์ตโฟน ทวิตเตอร์มีต้นกำเนิดจากการระดมความคิด ที่ถูกจัดขึ้นโดยบริษัทโอดีโอ ซึ่งเป็นบริษัทพอดแคสติง โดยดอร์ซีเป็นนักศึกษาจากมหาวิทยาลัยนิวยอร์ก ได้แนะนำความคิดการส่งบริการข้อความสั้น พูดคุยและสนทนากันภายในกลุ่มเล็ก ๆ (Times, 2010)

สถิติการใช้โซเชียลมีเดียของประเทศไทยถึงต้นปี 2566 มีผู้ใช้โซเชียลมีเดียในประเทศไทย 52.25 ล้านคน ในเดือนมกราคม 2566 โดย 52.3 เปอร์เซ็นต์ เป็นผู้หญิง และ 47.7 เปอร์เซ็นต์ เป็นผู้ชาย โดยมีการแบ่งเป็นแอพลิเคชั่น 8 แพลทฟอร์ม ดังนี้ (Thai Social Media Behavior Stat & Insight 2023)

- Facebook มีผู้ใช้ 48.10 ล้านคน
- YouTube มีผู้ใช้ 43.90 ล้านคน
- TikTok มีผู้ใช้ 40.28 ล้านคน
- Facebook Messenger มีผู้ใช้ 35.05 ล้านคน
- Instagram มีผู้ใช้ 17.35 ล้านคน
- Twitter มีผู้ใช้ 14.60 ล้านคน
- LinkedIn มี “สมาชิก” เป็นคนไทย 4.10 ล้านราย
- Snapchat มีผู้ใช้ 525.0 พันคน

ระบบสื่อสังคมออนไลน์ในแต่ละแพลทฟอร์ม จะมีการใช้งานที่แตกต่างกันออกไปเช่น เฟสบุ๊ค ส่วนใหญ่จะเป็นการโพสต์ข้อมูล ยูทูป จะเป็นแพลทฟอร์มที่แสดงเกี่ยวกับวีดีโอ ที่มีความยาว เพื่อความบันเทิง หรือข่าวสาร ติ๊กต๊อก จะเป็นแพลทฟอร์มที่แสดงเกี่ยวกับวีดีโอสั้น ในลักษณะที่เป็นภาพแนวตั้ง อินสตาแกรม เป็นแพลทฟอร์มที่แบ่งปันรูปภาพ โดยเป็นการโพสต์ที่ต้องมีรูปภาพด้วยเท่านั้น ทวิตเตอร์ เป็นแพลตฟอร์มที่เน้นการแบ่งปันข้อมูลในรูปแบบข้อความ ที่จะมี แฮชแท็ก เพื่อรวบรวมข้อมูลที่อยู่ในเรื่องเดียวกันก็ได้ โดยเน้นข้อมูลข่าวสารและมีการแบ่งปันอย่างรวดเร็ว

3 การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP)

การประมวลผลภาษาธรรมชาติ (NLP) เป็นเทคโนโลยีแมชชีนเลิร์นนิ่งที่ช่วยให้คอมพิวเตอร์สามารถตีความ จัดการ และทำความเข้าใจภาษามนุษย์ได้ องค์กรในปัจจุบันมีข้อมูลเสียงและข้อความจำนวนมากจากช่องทางการสื่อสารต่างๆ เช่น อีเมล ข้อความ ฟีดข่าวโซเชียลมีเดีย วิดีโอ

เสียง และอื่นๆ พวกเขาใช้ซอฟต์แวร์ NLP เพื่อประมวลผลข้อมูลนี้โดยอัตโนมัติ วิเคราะห์เจตนาหรือความเชื่อมั่นในข้อความ และตอบสนองการสื่อสารของมนุษย์แบบเรียลไทม์

การประมวลผลภาษาธรรมชาตินั้นช่วยให้อุปกรณ์คอมพิวเตอร์ต่าง ๆ สามารถสื่อสารกับมนุษย์ได้ด้วยการใช้งานภาษาของเครื่องเอง และดำเนินการทำงานต่าง ๆ ที่เกี่ยวข้องกับภาษาได้ ยกตัวอย่างเช่น NLP นั้น ช่วยให้อุปกรณ์และคอมพิวเตอร์สามารถอ่านอักขระภาษาปกติ หรือทำความเข้าใจและตีความคำพูดของมนุษย์ ไปจนถึงการวัดอารมณ์ ความรู้สึกที่แฝงอยู่ในข้อความเหล่านั้นและกลั่นกรองใจความหรือนัยยะที่สำคัญออกมาเพื่อใช้งานระบบที่ทันสมัยในปัจจุบัน สามารถวิเคราะห์ข้อมูลในปริมาณมหาศาลเกินกว่าขีดความสามารถของมนุษย์ โดยตัดข้อจำกัดเรื่องความเหน็ดเหนื่อยออกไป และสามารถทำงานด้วยความแม่นยำ คงเส้นคงวา และปราศจากอคติ การทำงานในปัจจุบัน มักต้องรับมือกับข้อมูลดิบจำนวนมหาศาล

เนื่องจากภาษาที่มนุษย์ใช้กันนั้น มีความซับซ้อนและหลากหลายอย่างยิ่ง เพราะมนุษย์มีวิธีการแสดงออกมากมายนับไม่ถ้วน ทั้งในด้านการสื่อสารด้วยคำพูดหรือข้อความที่เกิดขึ้นด้วยการเขียน นอกจากการมีภาษานับร้อย ๆพัน ๆ ภาษา ซึ่งต่างมีภาษาถิ่นแยกย่อยลงไปอีกนั้น ทุกภาษายังทวีความซับซ้อนยิ่งขึ้นไปอีกด้วยการมีชุดไวยากรณ์และโครงสร้างทางภาษาเฉพาะตัวของตนเอง รวมถึงคำ กลุ่มคำ และแม้แต่ศัพท์แสลงต่างๆ และเมื่อมนุษย์เราใช้ภาษาในการสื่อสารกันนั้น เรายังมักนิยมเขียนข้อความในรูปแบบย่อ ละเครื่องหมายวรรคตอนออกไป หรือแม้แต่การสะกดคำผิด ส่วนการสื่อสารด้วยวาจานั้นก็ยังมีประเด็นท้าทายของภาษาถิ่นและสำเนียงเฉพาะของแต่ละภูมิภาค แม้แต่ในภาษาเดียวกัน รวมถึงการพูดที่ไม่ชัดเจน อ้ำอึ้ง หรือใช้คำทับศัพท์แทรก (SAS Institute, 2020)

การวิเคราะห์เหมืองดัชนีถ้อยคำจากข้อมูลที่ได้มาจากการดึงข้อมูลผ่าน Twitter นั้นจะเป็นข้อความที่ได้จากการทวีต เหตุการณ์หรือรายละเอียดของเหตุการณ์อุบัติเหตุผ่านตัวอักษร ข้อมูลสื่อสังคมออนไลน์ Twitter เป็นการใช้งานเพื่อตอบสนองต่อความต้องการแบ่งปันความรู้สึกหรือการบรรยายข้อมูลต่อสาธารณะ ซึ่งสามารถแสดงให้เห็นถึงสภาพเหตุกาณ์หรือพื้นที่เกิดเหตุ การวิเคราะห์เหมืองดัชนีถ้อยคำแบ่งออกเป็น 2 ส่วนหลักได้แก่ การประมวลผลภาษาธรรมชาติ (Natural language processing: NLP) และการหากฎความสัมพันธ์ (Association Rule mining: ARM) (อสมาภรณ์ สิทธิ, 2561)

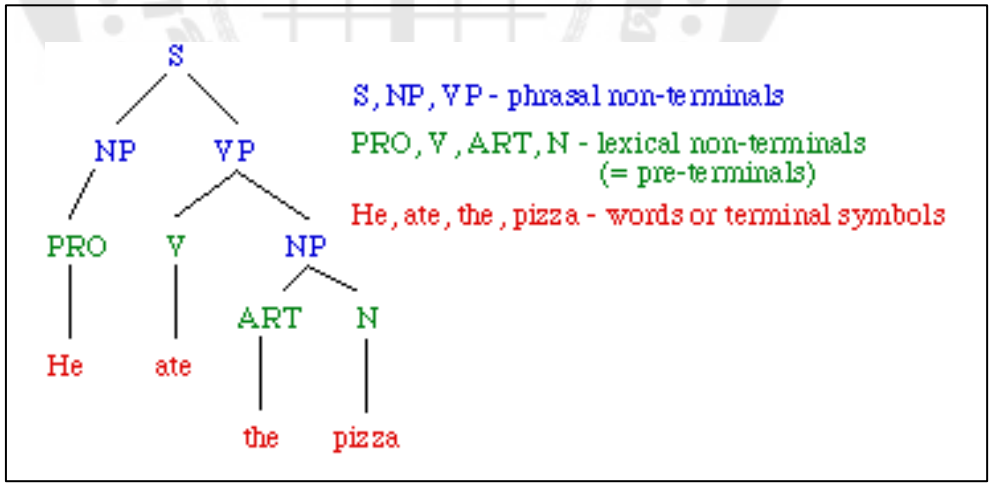
การทำเหมืองข้อความหรืออาจจะเรียกว่า การค้นหาความรู้ในฐานข้อมูลเอกสาร (Knowledge Discovery in Document Databases) เป็นเทคนิคเพื่อค้นหารูปแบบ (pattern) ของจากข้อความจำนวนมหาศาลโดยอัตโนมัติ โดยใช้ขั้นตอนวิธีจากวิชาสถิติ การเรียนรู้ของเครื่อง

และ การรู้จำแบบ หรือในอีกนิยามหนึ่ง การทำเหมืองข้อความ คือ กระบวนการที่กระทำกับข้อความ (โดยส่วนใหญ่จะมีจำนวนมาก) เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อความนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง หลักคณิตศาสตร์ หลักการประมวลเอกสาร (Document Processing) หลักการประมวลผลข้อความ (Text Processing)

3.1 การประมวลผลข้อความ (Text Processing)

การประมวลผลข้อความเป็นขั้นตอนการเตรียมข้อมูลสำหรับการวิเคราะห์เหมืองดัชนีถ้อยคำจากสื่อสังคมออนไลน์เพื่อให้ได้ข้อมูลที่เหมาะสมและมีความผิดพลาดน้อยที่สุด การประมวลผลข้อความ ประกอบด้วย 2 กระบวนการ ได้แก่ การเลือกคุณลักษณะของข้อมูลและการกรองข้อมูล ซึ่งมีรายละเอียดดังนี้ (นงนุช, ซาเฮาะ, & ตุ้มทอง, 2557)

การวิเคราะห์ในเชิงโครงสร้าง (Syntactic Analysis) เป็นการตรวจสอบโครงสร้างทางไวยากรณ์เกี่ยวกับการวาง ตำแหน่งของกลุ่มคำ ที่รวมกันเป็นประโยค ในกรณีที่ประโยคที่รับเข้ามาไม่ถูกต้องตามหลักไวยากรณ์ คอมพิวเตอร์ควรจะบอกได้ว่า เป็นประโยคที่ผิด โดยโครงสร้างของประโยค จะประกอบด้วย ภาคประธาน (Noun Phrase : NP) กับ ภาคแสดง (Verb Phrase : VP) เช่น ชื่อสถานที่หรือเหตุการณ์ ดังภาพประกอบ 5



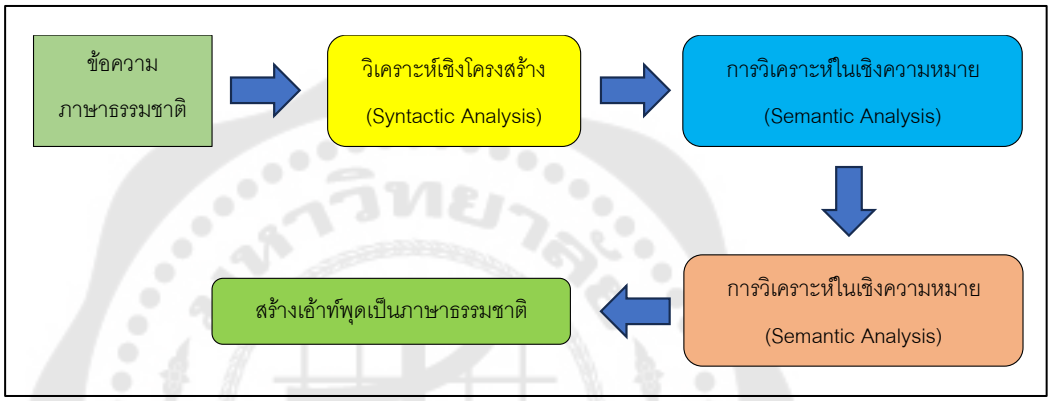
ภาพประกอบ 5 ลักษณะการตรวจสอบโครงสร้างทางไวยากรณ์

ที่มา : (นงนุช et al., 2557)

การวิเคราะห์ในเชิงความหมาย (Semantic Analysis) เป็นการตรวจสอบความถูกต้องในเชิงความหมายของประโยค โดยประโยคที่วางกลุ่มคำชนิดต่างๆตามโครงสร้างไวยากรณ์ จะมีความหมาย อย่างใดอย่างหนึ่งแน่นอนแต่ในบางครั้งประโยค ที่กำลังพิจารณาอาจจะเขียน ถูกต้อง

ตามหลักไวยากรณ์ แต่มีความหมายกำกวมหรือเป็นความหมายที่เป็น ไปไม่ได้ หรือไม่ให้ความหมายอะไรเลย (นงนุช et al., 2557)

การวิเคราะห์ในเชิงตีความ (Pragmatic Analysis) ประโยคที่เราพูดออกมาบางครั้งก็อาจจะไม่ได้มีความหมาย ตรงตาม ข้อความ นั้นๆซึ่งจะต้องตีความตามสถานการณ์ที่เกิดขึ้น โดยที่ทั้งผู้ส่ง ข่าวสารและผู้รับข่าวสารจะต้องอยู่ในสถานการณ์เดียวกัน (สำนักคอมพิวเตอร์ มหาวิทยาลัยทักษิณ, 2562)



ภาพประกอบ 6 โครงสร้างการวิเคราะห์ภาษาธรรมชาติ

ที่มา : กรมวุฒิ นงนุช และคณะ (2557)

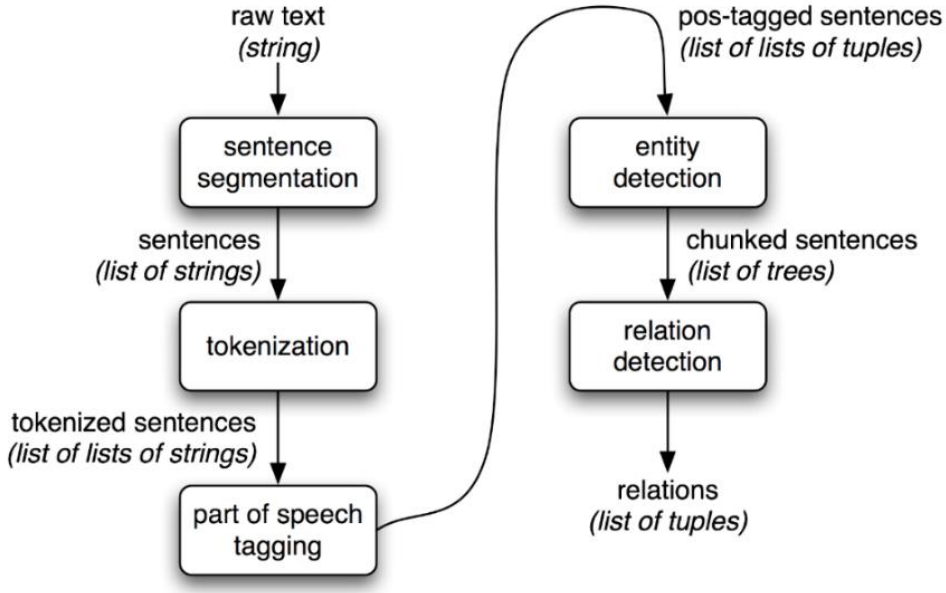
การประมวลผลภาษาธรรมชาติจากข้อมูลสื่อสังคมออนไลน์มีขั้นตอน ดังนี้

3.1.1 การเตรียมข้อมูล (Text Preprocessing phase)

การเตรียมข้อมูล คือ การที่เรานำข้อมูลมาเตรียมก่อนที่จะส่งข้อมูลดังกล่าวไปประมวลผลก่อน หรือพูดง่ายๆคือเราช่วยให้คอมพิวเตอร์ใช้งานข้อมูลของเราได้มีประสิทธิภาพมากขึ้น โดยเฉพาะในงานที่เกี่ยวข้องกับ NLP นั้นการเตรียมข้อมูลแทบจะเป็นหัวใจสำคัญของงานเลย เพราะว่าคอมพิวเตอร์ไม่สามารถเข้าใจภาษามนุษย์ได้เลยถ้าไม่ได้ถูกจัดการอย่างถูกต้องก่อน เหตุผลคือ คอมพิวเตอร์จัดการข้อมูลผ่านการทำงานของวงจรอิเลคทรอนิค ซึ่งถูกกำหนดโดยชุดคำสั่งที่ถูกแปลงจากภาษาคอมพิวเตอร์ซึ่งต้องมีมีความชัดเจนว่าต้องการให้ทำอะไร ขณะที่ภาษามนุษย์จะใช้การบรรยายสิ่งต่างๆ เพื่อสื่อสารระหว่างมนุษย์ มีความกำกวม และสามารถเปลี่ยนแปลงได้ตามบริบทที่แตกต่างกัน

3.1.2 การสกัดข้อมูล (Information Extraction) และการกรองข้อมูล (Data Filtration)

การสกัดคุณลักษณะของข้อความ (Text Extraction) โดยคัดเลือกเฉพาะส่วนที่ต้องการด้วยคำเดี่ยว พยางค์ วลี กลุ่มคำ หรือประโยค และกำจัดคำหยุด (Stopword) โดยตัดคำที่ไม่มีนัยสำคัญต่อเอกสารและไม่ทำให้ใจความของเอกสารเปลี่ยน เช่น คำสรรพนาม คำเชื่อม ชื่อเฉพาะ เช่น ชื่อถนน เป็นต้น (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557)



ภาพประกอบ 7 การสกัดข้อความที่ต้องการนำไปวิเคราะห์

ที่มา : (Steven Bird, 2019)

การกรองข้อมูล การกรองข้อมูลแบ่งออกเป็น 2 ประเภท ได้แก่ การกรองข้อมูลที่เป็นข้อความและการกรองข้อมูลเชิงตำแหน่ง ซึ่งการกรองข้อมูลข้อความเป็นการกลั่นกรองและแก้ไขข้อมูลคำผิดหรือบิดเบือนไปจากปกติ (Outlier data) เช่น ตัวสะกด โดยเฉพาะชื่อสถานที่หรือคำที่ระบุถึงตำแหน่งต่างๆ ให้ถูกต้อง ในส่วนการกรองข้อมูลเชิงตำแหน่งอาจมีความคลาดเคลื่อนที่เกิดมาจากอุปกรณ์ที่ใช้ในการระบุตำแหน่ง หรืออาจมีการระบุตำแหน่งที่นอกเหนือจากขอบเขตที่ศึกษา จึงต้องมีการกรองข้อมูลเฉพาะในพื้นที่ที่ทำการศึกษา (สิทธิ, 2561)

3.1.3 การทำกลุ่มข้อมูล (Word Cloud) และความถี่ข้อมูล (Frequency)

การจัดลำดับข้อมูล โดยวิเคราะห์จำนวนความถี่ของการเกิดขึ้นของข้อความนั้นๆ ให้อยู่ในรูปแบบถุงข้อความ (Bag of word) และนำมาจัดลำดับข้อมูล แสดงผลในรูปแบบของดัชนีถ้อยคำก้อนเมฆ (Tag cloud) แสดงขนาดตามจำนวนความถี่ของคำ ข้อความที่มีความถี่มากจะแสดงคำ

ที่มีขนาดใหญ่กว่าคำอื่น ๆ ซึ่งในหลาย ๆ วงการหรือในทางการวิเคราะห์ข้อมูล (Data Analytics) มักจะเอาไปใช้แสดงคำที่นิยมใช้ในบทความ หรือคำที่พูดถึงกันอยู่ในโลกโซเชียลได้ เป็นต้น และกราฟแสดงความถี่ของดัชนีถ้อยคำ (Frequency of word) (พิศิษฐ์ บวรเลิศสุธี, 2565) ดังภาพประกอบ 8

ภาพประกอบ 8 ข้อมูลที่อยู่ในรูปแบบถุงข้อความและดัชนีถ้อยคำก้อนเมฆ

ที่มา : (ANIKÓ KÁLMÁN, 2019)

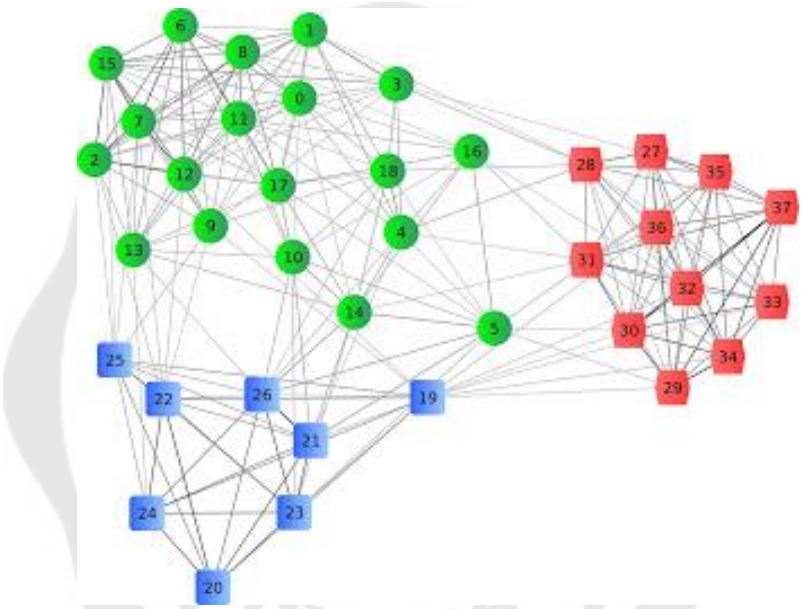
3.2 การหาความสัมพันธ์ของเหมืองดัชนีถ้อยคำ

3.2.1 การจัดกลุ่มข้อมูล (Data Clustering)

การจัดกลุ่มข้อมูลเป็นวิธีการจัดกลุ่มข้อมูลที่มีลักษณะเหมือนกันไว้ในกลุ่มเดียวกัน เป็นส่วนหลักของการการทำเหมืองข้อมูล การรู้จำแบบ เป็นการทำงานร่วมกันของอัลกอริทึมที่หลากหลายเพื่อแก้ปัญหาในการทำงาน ขั้นตอนวิธีที่ใช้ในการแบ่งกลุ่มจะอาศัยความเหมือน (similarity) หรือ ความใกล้ชิด (proximity) โดยจะแบ่งชุดข้อมูล ออกเป็นกลุ่ม (cluster) นำข้อมูลที่มีคุณลักษณะเหมือนกัน หรือคล้ายกันจัดไว้ในกลุ่มเดียวกัน

การจัดกลุ่มข้อมูลจะพิจารณาจากลักษณะการเกิดขึ้นของแต่ละชุดข้อมูลตามคุณลักษณะ (feature) เช่น คน สัตว์ สิ่งของ สถานที่ ภูมิทัศน์ กิจกรรมต่างๆ หรือขั้วความคิดเห็น (polarity) ข้อเสนอแนะและความรู้สึก ตามช่วงเวลาการแบ่งปันข้อมูลในแต่ละสถานที่ (สิทธิ, 2561)

การวิเคราะห์การจัดกลุ่มข้อมูล เป็นเทคนิคในกลุ่มของการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) จะทำการจัดกลุ่มข้อมูลที่ไม่เคยมีการจัดกลุ่มมาก่อนหน้า แต่จะแบ่งกลุ่มข้อมูลโดยพิจารณาจากลักษณะที่คล้ายกันของข้อมูล โดยจะนำข้อมูลที่มีลักษณะคล้ายกันมาอยู่กลุ่มเดียวกัน ส่วนข้อมูลที่มีลักษณะต่างออกไปก็ให้ไปอยู่อีกกลุ่มหนึ่ง แนวทางของการนำเทคนิคนี้ไปใช้ คือ จะไม่ใช่การหาผลลัพธ์ที่ต้องการวัดค่าความแม่นยำ หากแต่ต้องการหาความสัมพันธ์ของข้อมูลอีกรูปแบบหนึ่ง เช่น การจัดกลุ่มลูกค้าที่มีลักษณะความต้องการที่เหมือนกัน หรือจัดกลุ่มสินค้าที่มีคุณลักษณะเหมือนกัน (H, 2020) ดังภาพประกอบ 9



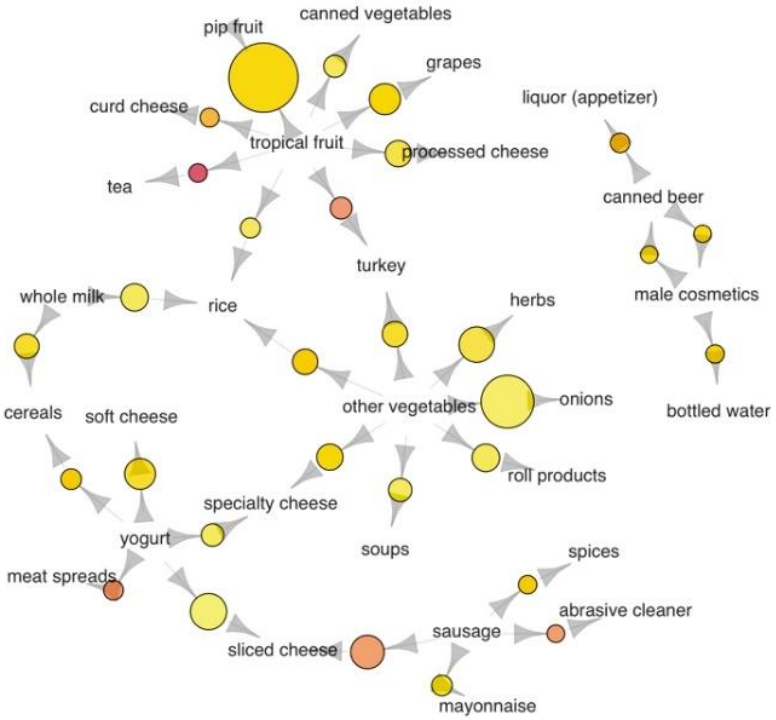
ภาพประกอบ 9 การจัดกลุ่มของข้อมูล

ที่มา : (Antoine Limasset, 2016)

3.2.2 กฎของการหาความสัมพันธ์ (Association Rule Mining: ARM)

การหากฎความสัมพันธ์(Association rules mining) เป็นเทคนิคหนึ่งในการทำเหมืองข้อมูลที่ค้นหาความสัมพันธ์ของเหตุการณ์หรือวัตถุที่เกิดขึ้นร่วมกันหรือพร้อมกันแล้วนำมาสร้างกฎขึ้นมาเพื่อที่จะทำนายเหตุการณ์หรือการเกิดขึ้นของวัตถุนั้นๆ ในอนาคตซึ่งการหากฎความสัมพันธ์นั้นสามารถนำไปใช้งานได้หลายรูปแบบ เช่น การวิเคราะห์พฤติกรรมการซื้อสินค้า หากลูกค้าซื้อสินค้า A แล้วมักจะซื้อสินค้า B ตามไปด้วย เป็นต้น และการเพิ่มประสิทธิภาพในการหากฎความสัมพันธ์ในรูปแบบต่าง ๆ เช่น การเพิ่มความเร็วในการหากฎความสัมพันธ์การเพิ่มค่าความถูกต้อง (Accuracy) หรือการประยุกต์ใช้การหากฎความสัมพันธ์กับข้อมูลในรูปแบบต่าง ๆ (Agrawal, 1993) ขั้นตอนการหาความสัมพันธ์ระหว่างคำนี้ประยุกต์กฎของการหาความสัมพันธ์

ของข้อมูลจากดัชนีถ้อยคำจะแยกกฎของความสัมพันธ์ออกจากกันโดยอัตโนมัติ โดยกำหนดชุดของดัชนีถ้อยคำสำคัญหลัก (K) จากข้อมูลที่ผ่านการกรองข้อมูลความถี่เชิงพื้นที่ มาจัดทำดัชนีคอลเล็กชันของดัชนีถ้อยคำ Ki = (k1,k2,…kn) เมื่อ n คือจำนวนประโยคทั้งหมดในเอกสาร และหาความสัมพันธ์ระหว่างดัชนีถ้อยคำต่างๆ (ki) ที่เกี่ยวข้องกับดัชนีถ้อยคำสำคัญหลัก (K) ในกรณีที่ ki มีความสัมพันธ์กับ K (Ki⊆K) มากนั้น กล่าวได้ว่าดัชนีถ้อยคำ ki มักจะพบอยู่กับดัชนีถ้อยคำสำคัญหลัก K บ่อยๆ หรือเป็นกลุ่มคำที่มีความคลายคลึงกันมา ดังภาพประกอบ 10 (คะอังกุ, 2555)



ภาพประกอบ 10 การหาความสัมพันธ์ของข้อมูล

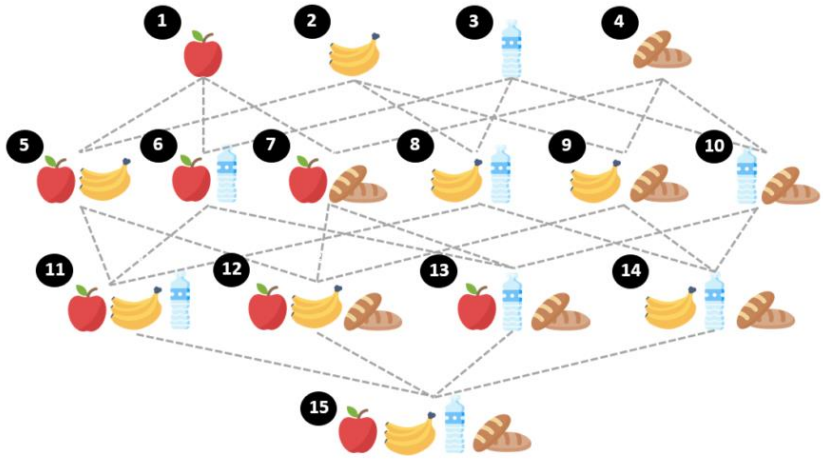
ที่มา : (Tanjil Ahmed, 2019)

3.3 อัลกอริทึมอพริโอริ (Apriori Algorithm)

การหากฎความสัมพันธ์ระหว่างข้อมูลด้วยอัลกอริทึมอพริโอริ เป็นอัลกอริทึมที่มีความนิยมในการใช้หาความสัมพันธ์ของข้อมูล โดยใช้หลักการค้นหาแบบวงกว้าง ซึ่งจะทำการสร้างและตรวจสอบ item sets ที่เกิดขึ้นบ่อยทีละชั้น การหากฎความสัมพันธ์ด้วยวิธี Apriori มี 2 ขั้นตอนใหญ่ๆ คือ การหา Frequent item sets คือการหารูปแบบข้อมูลที่เกิดขึ้นร่วมกันบ่อยๆในฐานข้อมูล หรือมีค่ามากกว่าค่าสนับสนุนต่ำสุด (Minimum support) ที่ผู้ใช้กำหนด โดยจะแบ่งเป็น 2 ขั้นตอนย่อย คือ การสร้างรูปแบบของเซตไอเทม (Itemset) ที่มีขนาดยาวมากขึ้นทีละ

หนึ่งขั้นไปเรื่อยๆ จนกระทั่งไม่เหลือเซตไอเทมที่จะสร้างอีกต่อไปทำให้ได้กลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยทั้งหมด และ การนับค่าการสนับสนุน (Support count) หลังจากที่สร้างรูปแบบของ item sets ได้แล้วนำมาทำการคำนวณค่า support ที่เกิดขึ้น โดยที่ Support คือจำนวณเปอร์เซ็นต์ที่พบ item sets ในฐานข้อมูล การหาความสัมพันธ์ของเหมืองดัชนีถ้อยคำมีดังนี้ (Agrawal, et al., 1993)

อพริโอริเป็นเทคนิคหนึ่งของ Data Mining คือการค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่เพื่อนำไปหารูปแบบที่เกิดขึ้นบ่อยๆ (frequent pattern) และใช้ในการวิเคราะห์ความสัมพันธ์หรือทำนายปรากฏการณ์ต่างๆ โดยการใช้ฐานข้อมูลประเภท Transaction Database (บัวทอง, 2557)



ภาพประกอบ 11 อัลกอริทึมอพริโอริ

ที่มา : (Kanyawee Pornsawangdee, 2021)

ผลลัพธ์ที่ได้เป็นกฎความสัมพันธ์ (Association Rule) สามารถเขียนได้ในรูปเซตของรายการที่เป็นเหตุ ไปสู่เซตของรายการที่เป็นผล ซึ่งมีรากฐานมาจากการวิเคราะห์ตะกร้าตลาด (Market Basket Analysis) เช่น ลูกค้าที่ซื้อแอปเปิ้ลส่วนใหญ่จะซื้อกล้วยด้วย ข้อมูลที่นำมาใช้จะอยู่ในรูปแบบ Nominal หรือ Ordinal เท่านั้น

การวิเคราะห์ตะกร้าตลาด เป็นรูปแบบที่ใช้เพื่อหากลุ่มสิ่งของที่น่าจะ ปรากฏร่วมกันในtransaction หนึ่งๆ ซึ่งมักเป็น transaction ณ จุดขาย ผลลัพธ์ที่ได้สามารถแสดงได้ด้วยกฎ ซึ่งบอกความเป็นไปได้ของการซื้อ สินค้าต่างๆร่วมกัน การวิเคราะห์ตะกร้าตลาด มีบทบาทสำคัญต่ออุตสาหกรรมการ ค้าปลีก (retail industry) ซึ่งใช้สารสนเทศ ศึกษาพฤติกรรมของลูกค้า จัดพื้นที่

ร้านค้า จัดวางสินค้าร่วมกันเพื่อส่งเสริมการขาย การวางแผนการส่งเสริมการขายและตั้งราคาผลิตภัณฑ์

การหากฎความสัมพันธ์ระหว่างข้อมูลเพื่อแบ่งกลุ่มข้อความด้วยอัลกอริทึมอพริโอริ (Apriori Algorithm) โดยจะพิจารณาจาก 3 ค่า ได้แก่

ค่า Support คือค่าสนับสนุนของ Frequent itemsets หากค่า minimum support ต่ำกว่าที่กำหนดก็จะตัด Itemset นั้นทิ้ง

ค่า Confidence คือค่าความเชื่อมั่นของความสัมพันธ์ของ Itemset หากเกิด A มักจะเกิด B กี่ % หากค่า Confidence = 100% แสดงว่า ถ้าหากมีชื่อถนน (A) ที่มีความสัมพันธ์กับอุบัติเหตุก็จะมีคำที่เกี่ยวข้องเกิดขึ้นนี้ในข้อความ (B)

ค่า Lift คือค่า บอกถึงความถูกต้องของกฎ แสดงถึงความขึ้นต่อกันของ Itemset หากค่า Lift > 1 แสดงว่า A ขึ้นต่อ B และหากค่า Lift < 1 แสดงว่า A ไม่ขึ้นต่อ B (A และ B เป็นอิสระต่อกัน)

การสร้างกฎด้วยอัลกอริทึมอพริโอริ โดยกำหนดให้พารามิเตอร์ (Parameter) minimum confidence = 0.08 และ lift > 1

ในงานวิจัยนี้สามารถนำมาหากฎความสัมพันธ์ที่ปรากฏร่วมกันของการเกิดอุบัติเหตุในเชิงพื้นที่

4. ระบบสารสนเทศภูมิศาสตร์ (Geographic Information System: GIS)

ระบบสารสนเทศภูมิศาสตร์ หมายถึงระบบคอมพิวเตอร์และโปรแกรมที่ออกแบบเพื่อสนับสนุนการนำเข้าข้อมูล การจัดการข้อมูล การเปลี่ยนแปลงข้อมูล การวิเคราะห์ข้อมูล และแสดงผลข้อมูล เพื่อสนับสนุนการตัดสินใจ (กลิ่นดาว, 2542)

อยู่เมือง (2540) อธิบายว่า ระบบสารสนเทศภูมิศาสตร์เป็นระบบข้อมูลข่าวสารที่เก็บไว้ในคอมพิวเตอร์ แต่สามารถแปลความหมายเชื่อมโยงกับสภาพภูมิศาสตร์อื่นๆ สภาพท้องที่ สภาพการทำงานของระบบสัมพันธ์กับสัดส่วนระยะทางและพื้นที่จริงบนแผนที่ ข้อแตกต่างระหว่าง GIS กับ MIS นั้นสามารถพิจารณาได้จากลักษณะของข้อมูล คือ ข้อมูลที่จัดเก็บใน GIS มีลักษณะเป็นข้อมูลเชิงพื้นที่ (Spatial Data) ที่แสดงในรูปของภาพ (graphic) แผนที่ (map) ที่เชื่อมโยงกับข้อมูลเชิงบรรยาย (Attribute Data) หรือฐานข้อมูล (Database)การเชื่อมโยงข้อมูลทั้งสองประเภทเข้าด้วยกัน จะทำให้ผู้ใช้สามารถที่จะแสดงข้อมูลทั้งสองประเภทได้พร้อมๆกัน ระบบสารสนเทศภูมิศาสตร์ประกอบด้วย

4.1 ข้อมูลในระบบสารสนเทศภูมิศาสตร์

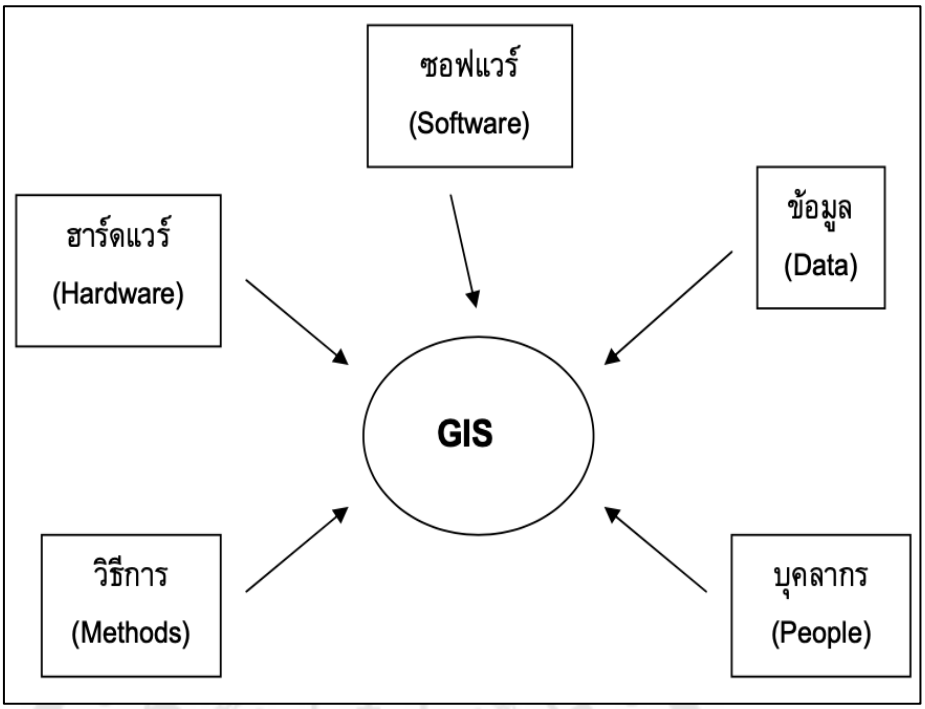
ข้อมูลเชิงพื้นที่ (Spatial data) คือ ข้อมูลที่แสดงถึงสิ่งต่างๆที่ปรากฏบนพื้นผิวโลกและอ้างอิงกับพิกัดทางภูมิศาสตร์ได้ สามารถจำแนกได้ 2 รูปแบบคือ ข้อมูลสารสนเทศเชิงกริด (Raster) และข้อมูลสารสนเทศเชิงตาราง (Vector)

1. ข้อมูลสารสนเทศเชิงกริด (Raster) คือ ข้อมูลที่มีโครงสร้างเป็นช่องสี่เหลี่ยม เรียกว่า จุดภาพ (Grid cell) เรียงต่อกันสามารถเก็บค่าได้ 1 ค่าข้อมูลราสเตอร์สามารถจัดเก็บข้อมูลเชิงพื้นที่โดยการแทนค่าข้อมูลจากพื้นที่จริงลงในจุดภาพซึ่งมีค่าตำแหน่งตามแนวแกนราบ (X,Y)

2.ข้อมูลสารสนเทศเชิงตาราง (Vector) คือข้อมูลที่สร้างขึ้นโดยอ้างอิงกับข้อมูลราสเตอร์เพื่อใช้แสดงถึงสิ่งต่างๆ ที่ปรากฏบนพื้นผิวโลก ประกอบด้วย ขนาดและทิศทาง มีรูปแบบแสดงได้ดังนี้

- จุด (point) ใช้แสดงข้อมูลที่เป็นลักษณะของตำแหน่งที่ตั้ง เช่น ที่ตั้งอาคารบ้านเรือน
- เส้น (Line) ใช้แสดงข้อมูลที่เป็นลักษณะของเส้น เช่น ถนน แม่น้ำ
- พื้นที่ (Polygon) ใช้แสดงข้อมูลที่เป็นลักษณะของพื้นที่ เช่น พื้นที่ขอบเขตการปกครอง (ทองสว่าง, 2520)

ข้อมูลเชิงบรรยาย (Attribute data) คือข้อมูลที่เกี่ยวกับคุณลักษณะเฉพาะของข้อมูลเชิงพื้นที่ข้อมูลเชิงบรรยาย มีลักษณะเป็นตัวอักษร (Character) หรือตัวเลข (Number) ซึ่งโดยทั่วไปจัดเก็บข้อมูลในรูปแบบของตาราง



ภาพประกอบ 12 องค์ประกอบของระบบสารสนเทศภูมิศาสตร์
ที่มา : ดัดแปลงจาก (จิรขจรกุล, 2555)

4.2 องค์ประกอบของระบบสารสนเทศภูมิศาสตร์

องค์ประกอบหลักของระบบ GIS จัดแบ่งออกเป็น 9 องค์ประกอบ ดังนี้

4.2.1 ฮาร์ดแวร์ (Hardware)

คือหน่วยประมวลผลกลาง (Center Processing Unit) เป็นเครื่องมือที่เป็นองค์ประกอบที่สนับสนุนกระบวนการสร้างฐานข้อมูลและจัดเก็บข้อมูลโดยมีการเชื่อมต่อกับจานบันทึกข้อมูลสำหรับเก็บข้อมูลหรือโปรแกรม เครื่องอ่านค่าพิกัด (Digitizer) สำหรับการแปลงข้อมูลแผนที่และเอกสารให้อยู่ในระบบดิจิตอล ไปเปิดยังเครื่องคอมพิวเตอร์ หรืออุปกรณ์อื่นๆ เพื่อแสดงผลลัพธ์จากการประมวลผลข้อมูล (ศูนย์วิจัยภูมิสารสนเทศเพื่อประเทศไทย)

4.2.2 ซอฟท์แวร์ (Software)

คือโปรแกรมหรือชุดคำสั่งที่สำหรับระบบสารสนเทศภูมิศาสตร์สั่งให้คอมพิวเตอร์ทำงานตามรูปแบบ เพื่อจัดการการนำเข้าข้อมูล ปรับแต่งข้อมูล เรียกค้นวิเคราะห์ และจำลองภาพ ตามวัตถุประสงค์ที่ต้องการ เช่น โปรแกรม Arcgis MapInfo ArcInfo QGis เป็นต้น (คีรีทอง, 2555)

4.2.3 วิธีการหรือขั้นตอนการปบัติงาน (Methodology)

คือขั้นตอนการทำงานในด้านระบบสารสนเทศภูมิศาสตร์เกี่ยวข้องกับวิธีการในการจัดเตรียมฐานข้อมูล การนำเข้าสู่ระบบ การจัดเก็บบันทึกข้อมูล การแสดงผลแผนที่ และการวิเคราะห์ข้อมูล ซึ่งเราเป็นผู้กำหนดให้เครื่องคอมพิวเตอร์จัดการกับข้อมูล (ศูนย์วิจัยภูมิสารสนเทศเพื่อประเทศไทย)

4.2.4 การนำเข้าข้อมูล (Data Input)

คือกระบวนการบันทึกข้อมูลเข้าสู่คอมพิวเตอร์ การสร้างฐานข้อมูลที่ละเอียด ถูกต้อง เป็นสิ่งสำคัญอย่างยิ่งในการปฏิบัติงานด้วยระบบสารสนเทศภูมิศาสตร์ซึ่งจำเป็นต้องมีการประเมินคุณภาพข้อมูล ที่จะนำเข้าสู่ระบบในเรื่องแหล่งที่มาของข้อมูล วิธีการสำรวจข้อมูลมาตราส่วนของแผนที่ ความถูกต้อง ความละเอียด พื้นที่ที่ข้อมูลครอบคลุมถึงและปีที่จัดทำข้อมูล เพื่อประเมินคุณภาพ และคักเลือกข้อมูลที่จะนำเข้าสู่ระบบฐานข้อมูล (ศูนย์วิจัยภูมิสารสนเทศเพื่อประเทศไทย)

4.2.5 การจัดการข้อมูล (Data Management)

คือกระบวนการจัดเก็บข้อมูลและแก้ไขข้อมูลจากระบบฐานข้อมูล วิธีการหรือเครื่องมือที่ช่วยในการจัดการฐานข้อมูล มีหลากหลายวิธี ข้อมูลดังกล่าจะถูกจัดเก็บโดยอ้างอิ่งจากค่าพิกัดทางภูมิศาสตร์ การจัดการฐานข้อมูล (Database) นิยมใช้โครงสร้างตามหลักการของฐานข้อมูลเชิงสัมพันธ์ (relational database) (ศูนย์วิจัยภูมิสารสนเทศเพื่อประเทศไทย)

4.2.6 การวิเคราะห์ข้อมูล (Data Analysis)

คือกระบวนการที่ปฏิบัติเพื่อให้ได้มาซึ่งสารสนเทศ ระบบสารสนเทศภูมิศาสตร์ การวิเคราะห์ข้อมูลโดยการนำเข้าข้อมูลเชิงพื้นที่ (Spatial data) หลายๆชั้นข้อมูลมาวางซ้อนทับกัน (Overlay) ตามเงื่อนไขที่ได้กำหนดไว้ เพื่อทำการวิเคราะห์และกำหนดเงื่อนไขต่างๆ ตามวัตถุประสงค์ หรือแบบจำลอง ซึ่งอาจเป็นการเรียกค้นข้อมูลอย่างง่าย โดยให้สัมพันธ์กับข้อมูลเชิงบรรยาย (Attribute data) เพื่อให้ได้ผลลัพธ์ที่ต้องการ

4.2.7 การแสดงผล (Data Display)

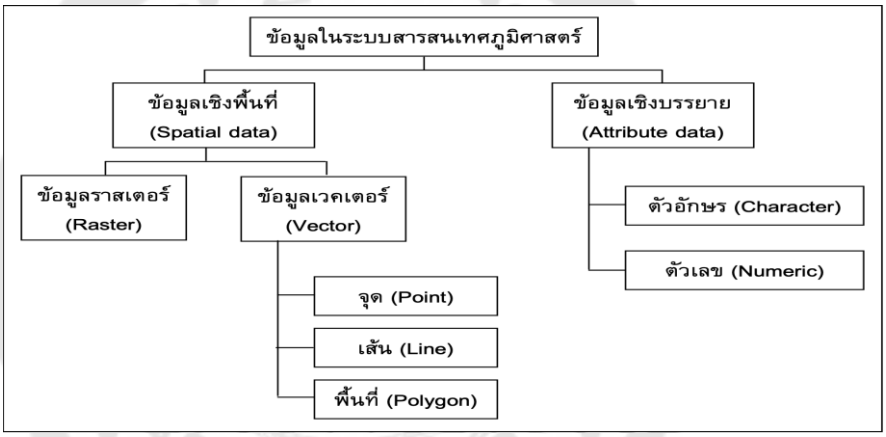
คือการนำเสนอผลต่อผู้ใช้ข้อมูล อาจอยู่ในรูปแบบของตัวเลข ข้อมูลรูปภาพ หรือรูปแบบตารางคำบรรยายแผนที่โดนให้ปรากฏทั้งบนสำเนาถาวร และภาพบนคอมพิวเตอร์ และสามารถแปลงข้อมูลเหล่านี้ไปสู่โปรแกรมอื่นในรูปแบบ แผนที่ แผนภูมิ หรือตาราง

4.2.8 บุคลากร (People)

คือผู้มีหน้าที่จัดให้องค์ประกอบทั้งหมดสามารถทำงานประสานกันจนได้ผลลัพธ์ระบบสารสนเทศภูมิศาสตร์ ในรูปแบบของข้อมูลและผลลัพธ์จากการวิเคราะห์ข้อมูล ต้องเป็นบุคคลที่มีความรู้ด้านคอมพิวเตอร์และภูมิศาสตร์ สามารถวิเคราะห์และออกแบบจัดทำแผนที่

4.2.9 ข้อมูล (Data)

คือข้อเท็จจริงต่างๆ ที่เกิดขึ้นหรือสถิติที่ใช้ในระบบสารสนเทศภูมิศาสตร์ จากแหล่งข้อมูลปฐมภูมิและทุติยภูมิ ถูกจัดเก็บในรูปแบบของฐานข้อมูลในคอมพิวเตอร์ และประมวลผลลัพธ์ออกมาเป็นส่วนประกอบที่เกี่ยวข้องและนำมาเชื่อมโยงกับระบบสารสนเทศภูมิศาสตร์ทั้งในรูปแบบแผนที่และข้อมูลสถิติที่เกี่ยวข้อง (สำนักงานเทคโนโลยีอวกาศและภูมิศาสตร์, 2552)



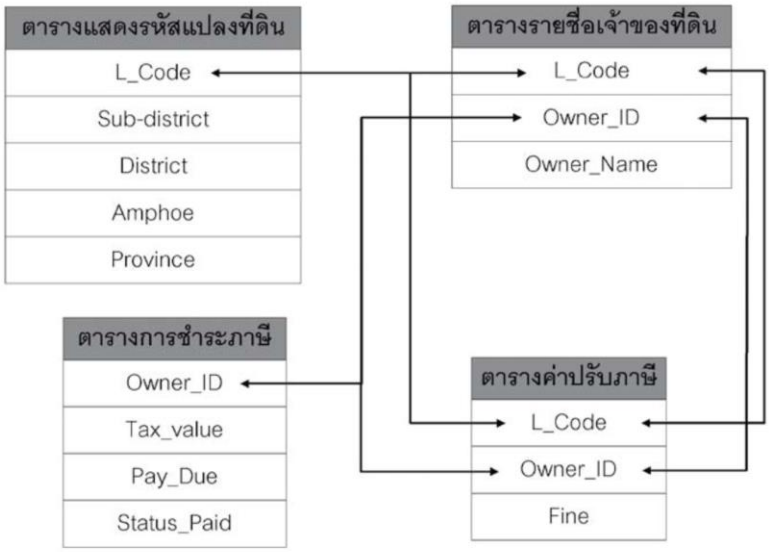
ภาพประกอบ 13 แสดงรายละเอียดข้อมูลสารสนเทศภูมิศาสตร์
ที่มา : (สำนักงานเทคโนโลยีอวกาศและภูมิศาสตร์, 2552)

4.3 ระบบฐานข้อมูล (Database)

ระบบฐานข้อมูลในระบบสารสนเทศภูมิศาสตร์รูปแบบการจัดเก็บข้อมูลแต่เดิมแบบแฟ้มข้อมูล ต่อมาได้มีการจัดข้อมูลให้อยู่ในรูปฐานข้อมูล เพื่อลดปัญหาการซ้ำซ้อนของข้อมูล (Data redundancy) และความขัดแย้งกันของข้อมูล อันมีสาเหตุจากการแก้ไขข้อมูล (Modification anomaly) การเพิ่มข้อมูล (Insertion anomaly) และการลบข้อมูล (Deletion anomaly) เป็นต้น ปัญหาเหล่านี้ ทำให้การวิเคราะห์ไม่ถูกต้องและขาดประสิทธิภาพ (ลวิโรจน์ วงศ์, 2557)

รูปแบบฐานข้อมูลแตกต่างรูปแบบแฟ้มข้อมูล โดยฐานข้อมูลเป็นการนำเอาข้อมูลต่างๆ ที่มีความสัมพันธ์กันซึ่งแต่เดิมจัดเก็บอยู่ในแต่ละแฟ้มข้อมูลมาจัดเก็บไว้ในที่เดียวกัน เช่น ข้อมูล

อาคาร ข้อมูลแปลงที่ดิน ข้อมูลการใช้ประโยชน์ที่ดิน ข้อมูลราคาประเมินภาษี ซึ่งเป็นข้อมูลด้านการจัดเก็บภาษีในเทศบาล และมีการจัดเก็บข้อมูลไว้ในแต่ละฝ่ายที่รับผิดชอบ เมื่อนำข้อมูลที่จัดเก็บในรูปแบบแฟ้มข้อมูล มาจัดเก็บอยู่ในแหล่งเดียวกันเป็นรูปแบบของฐานข้อมูล คือ ฐานข้อมูลของเทศบาล ส่งผลให้สามารถใช้ข้อมูลร่วมกัน และสามารถแก้ไขปัญหาที่เกิดจากการใช้งานข้อมูลแบบแฟ้มข้อมูลกรณีต่างๆ ได้ ข้อมูลที่จะนำมาจัดเก็บเป็นฐานข้อมูลได้นั้นจะต้องเป็นข้อมูลที่มีความสัมพันธ์กัน และสนับสนุนการดำเนินการอย่างใดอย่างหนึ่งขององค์กร ซึ่งจะเรียกว่า ระบบฐานข้อมูล (สำนักงานเทคโนโลยีอวกาศและภูมิศาสตร์, 2552)

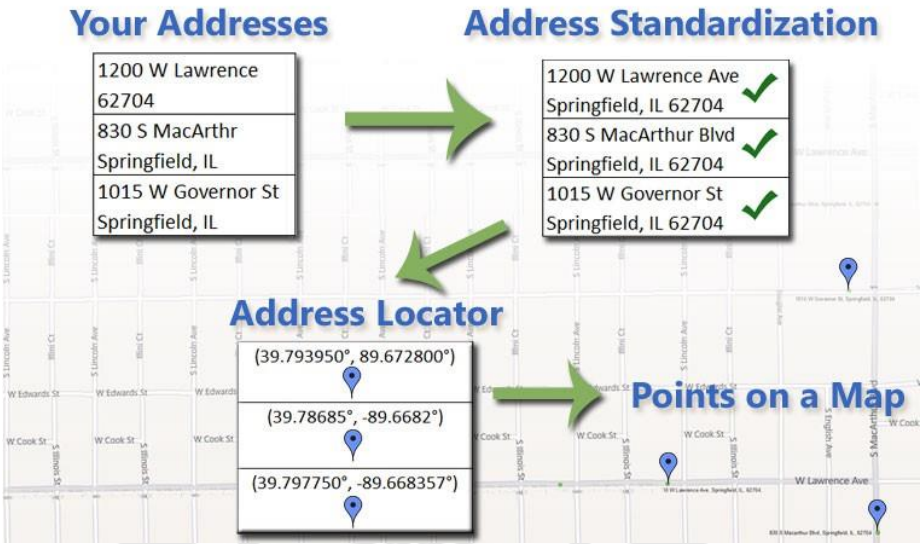


ภาพประกอบ 14 แบบจำลองฐานข้อมูล

ที่มา : (สำนักงานเทคโนโลยีอวกาศและภูมิศาสตร์, 2552)

4.4 กระบวนการแปลงข้อความเป็นพิกัดทางภูมิศาสตร์ (Geocoding)

กระบวนการแปลงข้อความเป็นพิกัดทางภูมิศาสตร์ (Geocoding) คือวิธีการเข้ารหัส แปลงที่อยู่ (address) ให้กลายเป็นตำแหน่ง พิกัดอ้างอิงบนแผนที่ เพื่อความสะดวกในการอ้างอิง ลดจำนวนข้อมูล ตัวอักษรในการระบุ บ้านเลขที่ ถนน หมู่บ้าน ตำบล ขณะเดียวกันปัจจุบันก็มีหลายผู้ให้บริการ ที่ให้บริการระบบการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์ และมีการใช้งานกันอยู่ในหลายรูปแบบ ดังภาพประกอบ 15 (ncm, 2018)



	geometry	address	road
0	POINT (100.69833 13.79070)	ถนนรามคำแหง, แขวงราษฎร์พัฒนา, เขตสะพานสูง, กรุ...	ถนนรามคำแหง
1	POINT (100.73998 13.72200)	ถนนกิ่งแก้ว, แขวงลาดกระบัง, เขตลาดกระบัง, กรุง...	ถนนกิ่งแก้ว
2	POINT (100.47702 13.87084)	ถนนรัตนาธิเบศร์, ตำบลไทรม้า, เทศบาลเมืองไทรม้า...	ถนนรัตนาธิเบศร์
3	POINT (100.29393 13.78636)	ถนนบรมราชชนนี, บางกระทึก, ขุนแก้ว, จังหวัดนครป...	ถนนบรมราชชนนี
4	POINT (100.63141 13.82270)	ถนนประเสริฐมนูกิจ, แขวงลาดพร้าว, เขตลาดพร้าว, ...	ถนนประเสริฐมนูกิจ

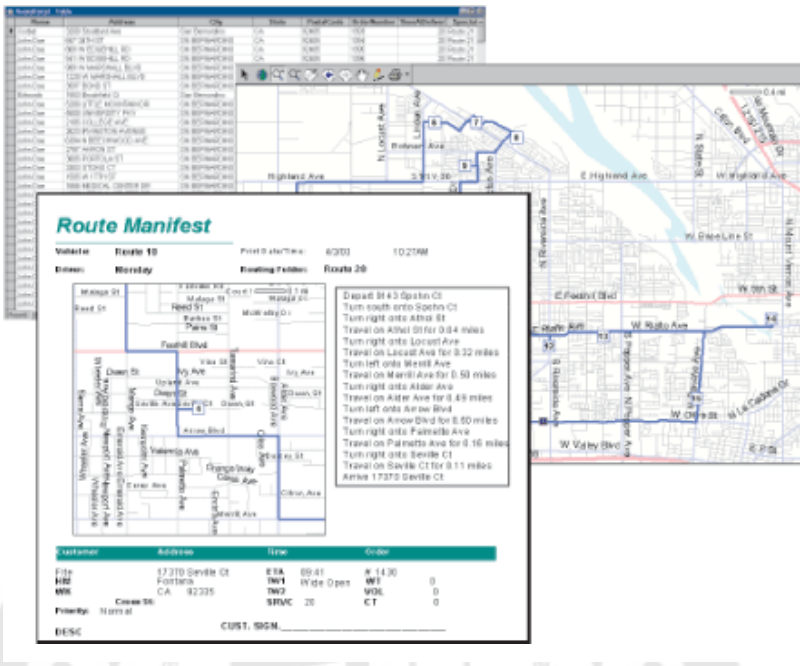
ภาพประกอบ 15 การแปลงข้อมูลที่อยู่ให้เป็นพิกัดทางภูมิศาสตร์

ที่มา : (Chakravarty, 2018)

Geocoding โดยทั่วไปประกอบด้วยขั้นตอนต่อไปนี้:

- การแยกวิเคราะห์ที่อยู่ (Address Parsing) : การแบ่งที่อยู่ออกเป็นส่วนประกอบแต่ละส่วน เช่น ชื่อถนน เมือง รัฐ และรหัสไปรษณีย์
- การทำให้เป็นมาตรฐาน (Standardization) : การรับประกันว่าส่วนประกอบของที่อยู่เป็นไปตามรูปแบบหรือโครงสร้างที่สม่ำเสมอ
- การจับคู่ (Matching) : การเปรียบเทียบที่อยู่ที่แยกวิเคราะห์และทำให้เป็นมาตรฐานกับฐานข้อมูลอ้างอิง เช่น บริการแปลงเป็นระบบพิกัด (Geocoding API Service) หรือชุดข้อมูล เพื่อหาผลลัพธ์ที่ดีที่สุด
- การแปลงข้อความเป็นพิกัด (Conversion) : การแปลงที่อยู่ที่จับคู่กันเป็นพิกัดทางภูมิศาสตร์ ซึ่งโดยทั่วไปแล้วจะเป็นละติจูดและลองจิจูด

- ผลลัพธ์ (Result) : การให้พิกัดที่ได้ ซึ่งสามารถนำไปใช้เพื่อวัตถุประสงค์ต่างๆ เช่น การทำแผนที่ การวิเคราะห์เชิงพื้นที่ หรือการรวมข้อมูลเพิ่มเติม



ภาพประกอบ 16 กระบวนการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์

ที่มา : (ESRI, 2021)

กระบวนการแปลงข้อมูลตำแหน่งที่เป็นข้อความหรือคำอธิบาย ให้เป็นพิกัดทางภูมิศาสตร์ ซึ่งสามารถนำไปพล็อตบนแผนที่ได้ โดยการแปลงชื่อถนน ที่อยู่ จุดสังเกต หรือข้อมูลตำแหน่งอื่นๆ ให้เป็นพิกัดเชิงพื้นที่ ซึ่งสามารถใช้สำหรับการทำแผนที่ การวิเคราะห์ และการสร้างภาพข้อมูล พิกัดทางภูมิศาสตร์ที่ได้จะเป็นข้อมูลตำแหน่งที่แม่นยำ สามารถระบุตำแหน่งเฉพาะบนพื้นผิวโลกได้อย่างแม่นยำ ช่วยให้สามารถรวมข้อมูลตำแหน่งเข้ากับเทคโนโลยีเชิงพื้นที่ต่างๆ ได้ เช่น ระบบสารสนเทศภูมิศาสตร์ (GIS) แพลตฟอร์มการทำแผนที่ และระบบนำทาง

Geocoding มีความสำคัญอย่างยิ่งในแอปพลิเคชันต่างๆ เช่น บริการตามตำแหน่ง การจัดการโลจิสติกส์และห่วงโซ่อุปทาน การวางแผนเมือง ระบบตอบสนองฉุกเฉิน และการวิเคราะห์โซเชียลมีเดีย เนื่องจากช่วยให้สามารถแสดงภาพและวิเคราะห์ข้อมูลเชิงพื้นที่ในลักษณะที่เข้าใจง่ายและมีบทบาทสำคัญในการทำให้ข้อมูลสามารถแสดงภาพและวิเคราะห์ได้เชิงพื้นที่ ช่วยให้เข้าใจทางภูมิศาสตร์ และสนับสนุนกระบวนการตัดสินใจในหลากหลายสาขา

การใช้งาน geocoding โดยการใช้โปรแกรมรหัสเปิดไพธอน (Open software Python) สามารถเลือกใช้ได้หลากหลาย ตัวอย่างหนึ่งที่สะดวกและง่าย คือการใช้ผ่าน Geopandas ซึ่งมี

ฟังก์ชัน geocode() รองรับงานด้านนี้นักพัฒนาสามารถแปลงที่อยู่ ไปเป็นข้อมูลรูปร่างเรขาคณิต (Polygon) ประเภทข้อมูลแบบจุด (Point) ที่มีพิกัดแผนที่อ้างอิงได้ทันที หรือประเภทของข้อมูลแบบเส้น (Line) ที่นำมาใช้ในงานวิจัยนี้ เพื่อแสดงผลถนนที่มีการเกิดอุบัติเหตุจากการนำข้อมูลบนสื่อสังคมออนไลน์มาวิเคราะห์

5. เอกสารและงานวิจัยที่เกี่ยวข้อง

อสมาภรณ์ สิทธิ (2561) ทำการวิเคราะห์ระบบเหมืองข้อมูลจากสื่อสังคมออนไลน์ เพื่อวิเคราะห์เหตุการจากการจราจร เนื่องจากระบบสื่อสังคมออนไลน์ในปัจจุบันมีการใช้งานเป็นจำนวนมากและข้อมูลมีความเป็นปัจจุบัน ทำให้รับรู้สถานการณ์ต่างๆ โดยใช้ข้อมูลการทวีตข้อความบนทวีตเตอร์จากบัญชี จส100 และ สวพ.91 เนื่องจากเป็นบัญชีที่มีผู้ใช้งานและรายงานสภาพการจราจรเป็นจำนวนมาก พื้นที่ศึกษาคือทางหลวงในกรุงเทพมหานคร มีวัตถุประสงค์เพื่อวิเคราะห์การประมาณค่าข้อมูลการจราจรจากการแชร์ข้อความตามข้อมูลที่ติดแท็กภูมิศาสตร์โดยใช้เทคนิคดัชนีเหมืองข้อความ การจัดประเภททวีตที่เกี่ยวข้องกับการเข้าชมใช้การเข้ารหัสแบบโทเค็น (Token) การวิเคราะห์ข้อความที่เกี่ยวข้องกับทวีต ด้วยข้อมูลที่ติดแท็ก (Tags) ที่เกี่ยวกับภูมิศาสตร์ วิธีการศึกษาใช้เทคนิคการทำเหมืองข้อมูลผ่านสื่อสังคมออนไลน์จาก Twitter เพื่อประเมินปริมาณการใช้งานโดยใช้เทคนิคเหมืองข้อความและรูปแบบการจำแนกประเภทของ Naive Bayes วิธีการที่นำเสนอมุ่งเน้นไปที่การระบุเหตุการณ์ที่เกี่ยวข้องกับการจราจรและวิเคราะห์ความสัมพันธ์ระหว่างรูปแบบการเข้าชมและคำทวีต กรณีศึกษาของแพลตฟอร์มของทวิตเตอร์จะแบ่งออกเป็น ขั้นตอนหลักคือ พัฒนาฐานข้อมูลและการวิเคราะห์ของคำภาษาไทยหรือข้อความเพื่อค้นหาความสัมพันธ์ของข้อมูลที่มีลักษณะการจราจร อัลกอริทึมประกอบด้วย : 1)การสืบค้นข้อมูลผ่าน twitter API ด้วยระบบฐานข้อมูล Copus 2) การประมวลผลข้อมูลและการเลือกข้อมูลข้อความ 3) การใช้เหมืองข้อความเพื่อกำหนดตำแหน่งของข้อความโดยใช้เทคนิคกฎการเชื่อมโยง ผลการศึกษาสำหรับการประเมินการจำแนกสภาพการจราจรจากข้อมูลจาก Twitter มีความแม่นยำมากคิดเป็นร้อยละ 97.5 ข้อมูลนี้สามารถเป็นตัวบ่งชี้ความนิยมช่วงเวลาเร่งด่วนการเข้าถึงการขนส่งและกิจกรรมพิเศษที่เกี่ยวข้องกับสถานที่เหล่านั้น สรุปได้ว่าพฤติกรรมของมนุษย์สามารถมีอิทธิพลต่อการใช้งาน twitter จากสื่อสังคมออนไลน์

Matci (2018) ทำการศึกษาระบบรหัสทางภูมิศาสตร์ (Geocoding) โดยใช้ข้อมูลจากภาษาธรรมชาติ (Natural Language Process) จากที่อยู่ของโรงเรียนระดับประถม ในเมืองเอสกิชีเฮียร์ ทางตะวันตกของประเทศตุรกี เพื่อนำมาปรับปรุงวิธีการระบุตำแหน่งจากระบบที่อยู่เดิม ให้มีความแม่นยำมากขึ้น ใช้เครื่องมือการระบุตำแหน่งด้วยระบบ Google Geocoding API และ

ArcGIS Geocoding API โดยนำมาทำงานร่วมกันแล้วจะทำให้มีค่าความคลาดเคลื่อนของพิกัดน้อยลง สามารถระบุตำแหน่งของที่อยู่ได้ชัดเจนมากขึ้น ผลจากการศึกษาเมื่อนำระบบ API ทั้ง 2 ระบบมาทำงานร่วมกันสามารถเพิ่มความแม่นยำของพิกัดได้สูงขึ้น50% และกระบวนการสร้างมาตรฐานช่วยปรับปรุงการระบุพิกัดที่ถูกต้องจาก 64.4% เป็น 99.1%

Shekhar, Setty, and Mudenagudi (2016) ทำการศึกษาปัญหาเกี่ยวกับการจราจรในเมืองที่มีประชากรหนาแน่นโดยใช้ข้อมูลเรียลไทม์จากระบบสื่อสังคมออนไลน์ตามผู้ใช้งานระบบคมนาคมในเมืองส่วนใหญ่จะมีการโพสต์อัปเดตเกี่ยวกับการจราจรบนสื่อสังคมออนไลน์ในรูปแบบทวีตหรือโพสต์บนเฟซบุ๊ก ด้วยข้อมูลจำนวนมากของปัญหาการจราจรจึงทำการรวบรวมระบุและสร้างตัวจำแนกความรู้สึกเพื่อตรวจสอบอารมณ์ของผู้เดินทางตลอดเวลา ความรู้นี้ใช้ในการวิเคราะห์และคาดการณ์รูปแบบการจราจรในสถานที่ที่กำหนด นอกจากนี้เรายังระบุสาเหตุที่เป็นไปได้ของการจราจรติดขัดในพื้นที่ใดพื้นที่หนึ่งโดยการวิเคราะห์ข้อมูลในอดีตที่รวบรวมไว้ จากการทำงานของเราเราสามารถนำเสนอแนวทางที่ไม่ถูกตรวจสอบประหยัดและเป็นทางเลือกอื่นสำหรับวิธีการตรวจสอบการจราจรที่ติดขัด

Aghajani, Dezfoulian, Arjroody, and Rezaei (2017) ทำการประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์เพื่อระบุพื้นที่และเวลารูปแบบของอุบัติเหตุบนท้องถนนโดยใช้สถิติเชิงพื้นที่แสดงผลในรูปแบบแผนที่ความหนาแน่นของจุดที่เกิดอุบัติเหตุ เปรียบเทียบจำนวนอุบัติเหตุในแต่ละรูปแบบของภูมิประเทศและนำมาเปรียบเทียบกับปริมาณน้ำฝนจะเห็นได้ว่ามีความสัมพันธ์กัน คือจำนวนอุบัติเหตุในภาคตะวันออกเฉียงเหนือ ของจังหวัดที่มีสภาพอากาศเป็นภูเขาและมีฝนตกมากกว่าถึงแม้ว่าจำนวณครั้งขออุบัติเหตุในภาคตะวันออกเฉียงเหนือของจังหวัดจะมีมาก แต่การบาดเจ็บและการเสียชีวิตจะน้อยกว่าเมื่อเทียบกับถนนทางตะวันตกเฉียงเหนือ อาจจะสรุปได้ว่าในถนนทางตะวันตกเฉียงเหนือแม้จะมีการจราจรน้อย แต่จำนวนผู้เสียชีวิตก็มากขึ้นอาจเป็นเพราะปัจจัยต่างๆเช่นการออกแบบส้นทางคมนาคมยังขาดปัจจัยด้านความปลอดภัย

ทวี อุทัยเศรษฐวัฒน์ (2550)กล่าวว่าการประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์ในการหาจุดอันตรายบนทางพิเศษ โดยใช้การหาจุดอันตรายด้วยวิธีอัตราการเกิดอุบัติเหตุ(Accident Rate Method) ในการจัดลำดับของจุดอันตรายบนทางพิเศษเฉลิมมหานคร (ระบบทางด่วนขั้นที่ 1) ทางพิเศษศรีรัช (ระบบทางด่วนขั้นที่ 1) และทางพิเศษฉลองรัช (ทางด่วนสายรามอินทรา - อาจณรงค์) และใช้ข้อมูลอุบัติเหตุและปริมาณจราจรและลักษณะทางกายภาพถนน จากการวิเคราะห์ผู้วิจัยสรุปว่าอุบัติเหตุส่วนใหญ่จะเกิดบริเวณทางโค้ง ทางแยกต่างระดับที่มีรัศมีแคบ และมีความลาดชัน

ชิษณุ and ชัยพล (2550) ศึกษาและพิสูจน์หาจุดอันตรายบนถนน โดยใช้วิธีอัตราการเกิดอุบัติเหตุวิกฤต (Critical Crash Rate Method) ข้อมูล ที่ใช้เป็นข้อมูลสถิติการเกิดอุบัติเหตุและปริมาณจราจรในช่วงเวลาปี พ.ศ. 2545 - 2547 บนถนนรามอินทราและถนนแจ้งวัฒนะ จากกรมทางหลวงโดย ผู้วิจัยได้ทำการศึกษาที่ระยะทาง 100 เมตร ทุกช่วง ถนน หลังจากการวิเคราะห์แล้ว ผู้วิจัยได้นำข้อมูลมาเปรียบเทียบเพื่อจัดลำดับความสำคัญ โดยคำนวณหา ค่าอัตราส่วนระหว่าง Crash Rate และ Critical Crash Rate (DANGEROUS FACTOR : DF) ซึ่งจุดใดที่มี ค่า DF มากกว่า 1.0 จะถูกมองว่าเป็น จุดอันตราย

Karimi, Sharker, and Asavasuthirakul (2011) ทำการศึกษากระบวนการทำงานของระบบกระบวนการแปลงข้อความเป็นพิกัดทางภูมิศาสตร์ (Geocoding) เพื่อแนะนำกระบวนการที่เหมาะสมที่สุดในการทำงานของระบบ ปัจจุบันบริการจำนวนมากที่สามารถระบุรหัสทางภูมิศาสตร์ได้มีให้บริการสำหรับนักวิทยาศาสตร์และนักวิจัยโดเมนนักพัฒนาซอฟต์แวร์และผู้ใช้ปลายทาง ด้วยเหตุผลหลายประการรวมถึงคุณภาพของฐานข้อมูลอ้างอิงและเทคนิคการแก้ไขที่อยู่ที่ระบุทางภูมิศาสตร์โดยบริการที่แตกต่างกันมักไม่ส่งผลให้อยู่ในตำแหน่งเดียวกัน เมื่อพิจารณาว่ามีบริการ geocoding ที่มีอยู่มากมายและสามารถเข้าถึงได้และบริการ geocoding แต่ละแห่งอาจใช้ฐานข้อมูลอ้างอิงและเทคโนโลยีการแก้ไขที่แตกต่างกันการเลือกบริการ geocoding ที่เหมาะสมซึ่งตรงตามความต้องการของแอปพลิเคชันหรือผู้ใช้เป็นงานที่ท้าทาย โดยเฉพาะอย่างยิ่งสำหรับบริการ geocoding ออนไลน์ซึ่งมักใช้เป็นกล่องดำและไม่ได้ให้ความรู้เกี่ยวกับฐานข้อมูลอ้างอิงและเทคนิคการแก้ไขที่ใช้ ในบทความนี้เรานำเสนออัลกอริธึมผู้แนะนำการเข้ารหัสทางภูมิศาสตร์ที่สามารถแนะนำบริการระบุพิกัดทางภูมิศาสตร์ออนไลน์ที่ดีที่สุดโดยการตระหนักถึงคุณลักษณะ (ความแม่นยำของตำแหน่งและอัตราการจับคู่) ของบริการและความต้องการของผู้ใช้และ / หรือแอปพลิเคชัน อัลกอริทึมได้รับการจำลองและวิเคราะห์โดยใช้บริการระบุพิกัดทางภูมิศาสตร์ออนไลน์ยอดนิยม 6 รายการสำหรับประเภทที่อยู่ที่แตกต่างกัน (เกษตรกรรมเชิงพาณิชย์อุตสาหกรรมที่อยู่อาศัย) และความต้องการ (อัตราการจับคู่ความแม่นยำของตำแหน่ง)

Karagoz (2016) ศึกษาการแยกข้อมูลตำแหน่งจากข้อมูลเครือข่ายสังคมที่มาจากฝูงชนด้วยผู้ใช้หลายล้านคนทั่วโลกข้อมูลโซเชียลมีเดียที่มาจากฝูงชนจึงเป็นแหล่งข้อมูลที่มีค่าสำหรับเหตุการณ์ที่เกิดขึ้นทั่วโลก โดยเฉพาะอย่างยิ่งไมโครบล็อกซึ่งเป็นเครือข่ายโซเชียลที่บังคับใช้ข้อความสั้น ๆ ได้รับความนิยมสูงเนื่องจากความพร้อมใช้งานในฐานะแอปพลิเคชันมือถือและการใช้งานจริงของข้อความสั้น การประมาณตำแหน่งของเหตุการณ์ที่ตรวจพบโดยการติดตามโพสต์ในไมโครบล็อกเป็นแรงจูงใจของการศึกษาล่าสุดจำนวนมาก การดึงข้อมูลสถานที่และการประมาณ

สถานที่จัดงานเป็นงานที่ท้าทายในการรักษาการรับรู้สถานการณ์ที่น่าพอใจโดยเฉพาะกรณีฉุกเฉิน เช่นไฟไหม้หรืออุบัติเหตุจราจร ปัจจุบัน Twitter เป็นหนึ่งในแพลตฟอร์ม microblogging ที่ได้รับความนิยมมากที่สุดและมีงานวิจัยล่าสุดที่มุ่งเป้าไปที่การตรวจจับเหตุการณ์ใหม่ ๆ ทางออนไลน์โดยติดตามทวีต ในการวิเคราะห์เหตุการณ์โดยทั่วไปแล้วนักวิจัยจะมุ่งเน้นไปที่คุณลักษณะเชิงพื้นที่ชั่วคราวของโพสต์ คุณสมบัติชั่วคราวแสดงถึงเวลาและลำดับของโพสต์ในขณะที่คุณสมบัติเชิงพื้นที่มีประโยชน์สำหรับการแยกตำแหน่งหรือการประมาณค่าสำหรับการจดจำชื่อบุคคลและการประมาณตำแหน่งจากไมโครบล็อก

Karimi, Durcik, and Rasdorf (2004) ศึกษาฐานข้อมูลเชิงพื้นที่มีข้อมูลที่เข้ารหัสทางภูมิศาสตร์ ข้อมูล Geocoded มีบทบาทสำคัญในการใช้งานด้านวิศวกรจำนวนมากเช่นการขนส่งและการศึกษาสภาพแวดล้อมที่ใช้ระบบข้อมูลภูมิสารสนเทศ (GIS) สำหรับการสร้างแบบจำลองและการวิเคราะห์เชิงพื้นที่เนื่องจากมีข้อมูลเชิงพื้นที่ (เช่นละติจูดและลองจิจูด) เกี่ยวกับวัตถุ ข้อมูลที่ GIS สร้างขึ้นได้รับผลกระทบจากคุณภาพของข้อมูลพิกัดภูมิศาสตร์ (เช่นพิกัด) ที่จัดเก็บในฐานข้อมูล ในการตัดสินใจอย่างเหมาะสมและมีเหตุผลโดยใช้ข้อมูล geocoded สิ่งสำคัญคือต้องยกเลิกการเข้าใจแหล่งที่มาของความไม่แน่นอนใน geocoding แหล่งที่มาหลัก ๆ ของความไม่แน่นอนใน geocoding มีอยู่สองแหล่งที่มาที่เกี่ยวข้องกับฐานข้อมูลที่ใช้เป็นชุดข้อมูลอ้างอิงสำหรับวัตถุ geocode และอีกแหล่งหนึ่งที่เกี่ยวข้องกับเทคนิคการแก้ไขที่ใช้ ปัจจัยต่างๆเช่นความสมบูรณ์ความถูกต้องความสอดคล้องสกุลเงินและความถูกต้องของข้อมูลในฐานข้อมูลอ้างอิงมีส่วนทำให้เกิดความไม่แน่นอนของอดีตในขณะที่ตรรกะและสมมติฐานเฉพาะที่ใช้ในเทคนิคการแยกขั้วมีส่วนช่วยในเรื่องหลัง จุดประสงค์หลักของบทความนี้คือเพื่อทำความเข้าใจความไม่แน่นอนที่เกี่ยวข้องกับเทคนิคการแก้ไขที่ใช้สำหรับ geocoding ในการดำเนินการดังกล่าวได้มีการใช้และทดสอบอัลกอริทึมการระบุพิกัดทางภูมิศาสตร์ 3 รายการและเปรียบเทียบผลลัพธ์กับข้อมูลที่รวบรวมโดย Global Positioning System (GPS) ผลการเปรียบเทียบโดยรวมไม่พบความแตกต่างอย่างมีนัยสำคัญระหว่างอัลกอริทึมทั้งสามประเภท

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการประมวลผลภาษาธรรมชาติเชิงพื้นที่ พบว่างานวิจัยส่วนใหญ่ ใช้วิธีการทางระบบสารสนเทศภูมิศาสตร์เพื่อการวิเคราะห์ความหนาแน่นเชิงพื้นที่ของการเกิดอุบัติเหตุเพียงอย่างเดียว งานวิจัยนี้จึงได้นำการประมวลผลภาษาธรรมชาติมาวิเคราะห์ข้อมูลสื่อสังคมออนไลน์และประมาณตำแหน่งเชิงพื้นที่เพื่อประมาณลักษณะการเกิดอุบัติเหตุในเชิงพื้นที่ ด้วยวิธีระบุพิกัดทางภูมิศาสตร์ (Geocoding)

บทที่3
วิธีดำเนินงานวิจัย

ในการวิจัยการประมวลผลภาษาธรรมชาติจากข้อมูลสื่อสังคมออนไลน์เชิงพื้นที่เพื่อคาดการณ์อุบัติเหตุ มีขั้นตอนการวิจัย ดังนี้

1 แผนการดำเนินงาน

2 อุปกรณ์และเครื่องมือที่ใช้ในงานวิจัย

3. ข้อมูลที่ใช้ในการศึกษา

4. การสืบค้นข้อมูลและการจัดการฐานข้อมูล

5. การประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ

6. การหาความสัมพันธ์ระหว่างข้อมูลจากการประมวลผลภาษาธรรมชาติ

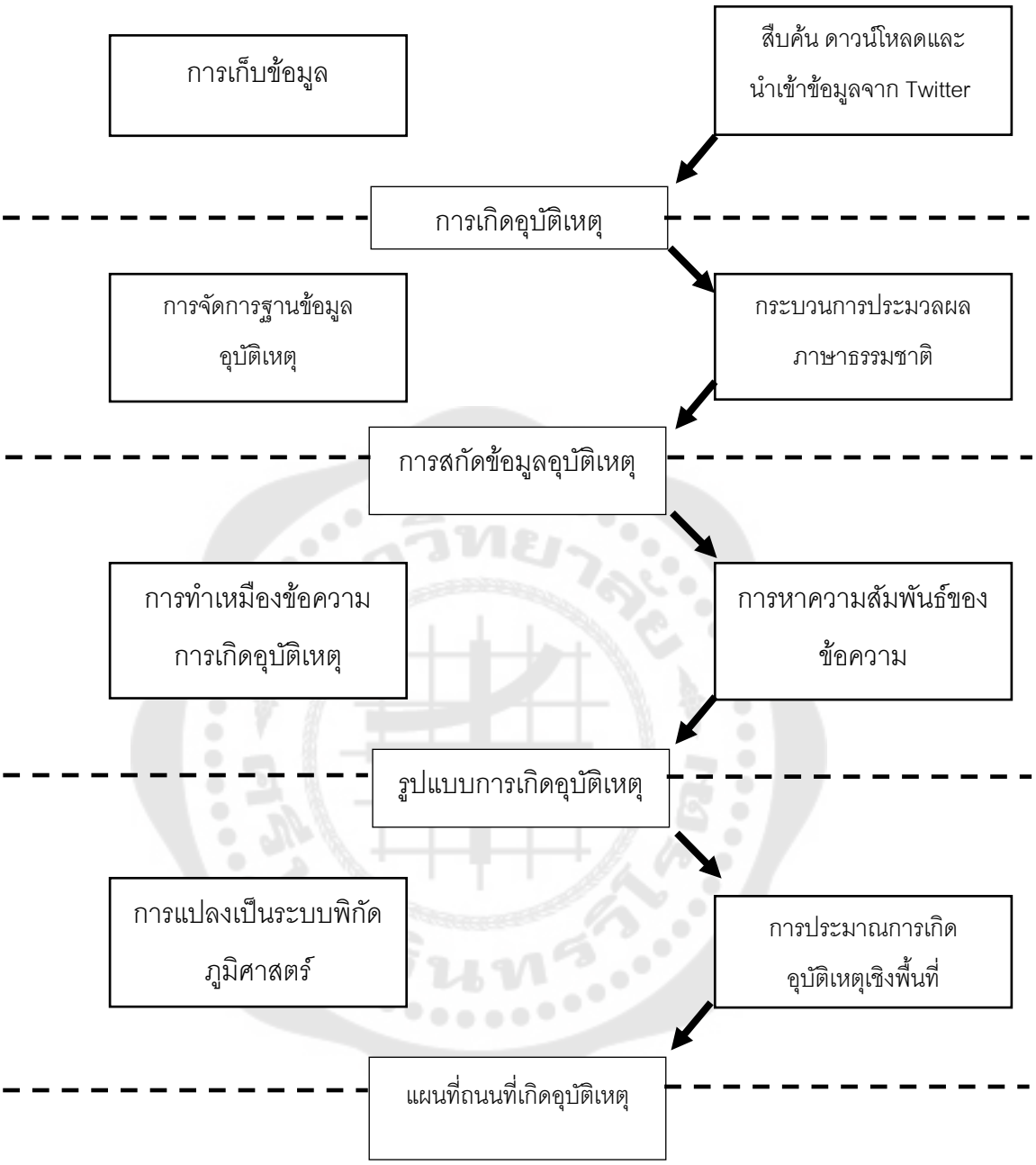
7. กระบวนการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์

8. การจัดทำแผนที่ความเสี่ยงจากการเกิดอุบัติเหตุ ตามช่วงเวลา

9. การตรวจสอบความสอดคล้องกับข้อมูลสถิติการเกิดอุบัติเหต

1. แผนการดำเนินงาน

การดำเนินงานในวิจัยนี้ใช้เทคนิคการวิเคราะห์เชิงตำแหน่งจากสื่อสังคมออนไลน์โดยวิธีการประมวลผลภาษาธรรมชาติและการทำเหมืองข้อมูลซึ่งเป็นกระบวนการหาความสัมพันธ์ของข้อมูลจากการประมวลผลภาษาธรรมชาติ โดยมีกระบวนการดังนี้



ภาพประกอบ 17 แผนการดำเนินงาน

2. อุปกรณ์และเครื่องมือที่ใช้ในงานวิจัย

2.1 คุณสมบัติเครื่องคอมพิวเตอร์ที่ใช้ในการพัฒนา

2.1.1 เครื่องคอมพิวเตอร์เป็นฮาร์ดแวร์ทหน้าที่เป็นเครื่องคอมพิวเตอร์แม่ข่าย (Server) สำหรับการเขียนโปรแกรมภาษา Python สำหรับการประมวลผลภาษาธรรมชาติและการหาความสัมพันธ์ระหว่างข้อความโดยมีคุณสมบัติดังนี้

2.1.1.1 หน่วยประมวลผลกลาง (CPU) แบบ 64 bit ความเร็ว 3.2 GHz

2.1.1.2 หน่วยความจำหลัก (RAM) ชนิด LPDDR4X ขนาด 8 GB

2.1.1.3 พื้นที่จัดเก็บข้อมูล (Hard Disk) ความจุ 256 GB จำนวน 1 หน่วย

2.1.1.4 ระบบปฏิบัติการ MacOS

2.2 โปรแกรม **(Software)** ที่ใช้ในการพัฒนา

2.2.1 โปรแกรม Node Red สำหรับการพัฒนาระบบการดึงข้อมูลผ่าน Twitter API

2.2.2 โปรแกรม Sequel Pro สำหรับการจัดการฐานข้อมูล MySQL

2.2.3 โปรแกรม Jupyter Notebook สำหรับการพัฒนาการประมวลผลภาษาธรรมชาติและการหาความสัมพันธ์ระหว่างข้อความและกระบวนการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์

2.2.4 โปรแกรมคอมพิวเตอร์ ArcGIS Version 10.8 โดยเป็นโปรแกรมหลักที่นำมาใช้งาน เพื่อวิเคราะห์และแสดงผลในขั้นตอนการวิเคราะห์ด้วยระบบสารสนเทศภูมิศาสตร์ และแสดงผลออกมาในรูปแบบของแผนที่

2.3 เทคโนโลยีและภาษาคอมพิวเตอร์ที่ใช้ในการพัฒนา

2.3.1 ระบบ API (Application Program Interface) สำหรับการช่วยให้ซอฟต์แวร์สื่อสารและเชื่อมต่อกัน

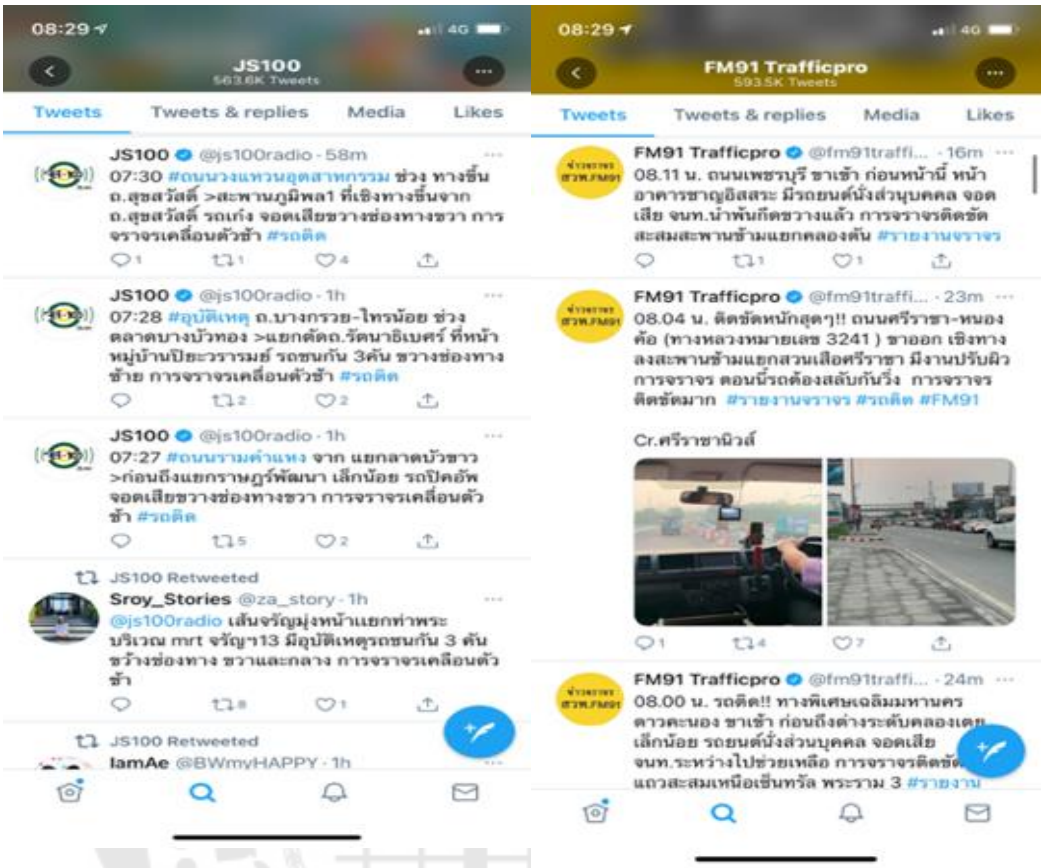
2.3.2 MySQL สำหรับเป็นระบบจัดการฐานข้อมูลเชิงสัมพันธ์ โดยใช้ภาษา SQL (Structured Query Language) สำหรับการจัดเก็บฐานข้อมูลและพัฒนาระบบฐานข้อมูล

2.3.3 ภาษา Python สำหรับการประยุกต์ใช้ในการพัฒนาระบบประมวลผลภาษาธรรมชาติ

2.3.4 การวิเคราะห์หาความสัมพันธ์ระหว่างข้อความด้วย Apriori Algorithm

3. ข้อมูลที่ใช้ในการศึกษา

ข้อมูลที่ใช้ในการศึกษาครั้งนี้ ได้ทำการทบทวนวรรณกรรมและเอกสารที่เกี่ยวข้องในเรื่องเกี่ยวกับการประมวลผลภาษาธรรมชาติจากสื่อสังคมออนไลน์เชิงพื้นที่ในการคาดการณ์อุบัติเหตุ โดยพื้นที่การศึกษาคือกรุงเทพมหานคร ใช้ข้อมูลการรายงานอุบัติเหตุจากสื่อสังคมออนไลน์ Twitter จากการทวีต ข้อความ (Tweet) หรือการทวีตข้อความนั้นซ้ำ (Retweet) ของบัญชี จส100 และ สวพ91 เนื่องจากเป็นบัญชีที่มีความน่าเชื่อถือของข้อมูล โดยมีการทวิตข้อความผ่านบัญชีผู้ใช้งานเป็นจำนวนมาก จึงเลือกใช้ 2 บัญชีนี้มาทำฐานข้อมูล ดังภาพประกอบ 18



ภาพประกอบ 18 ระบบสื่อสังคมออนไลน์ Twitter บัญชี JS100radio และ fm91trafficpro

ตาราง 5 ข้อมูลที่ใช้ในการวิจัย

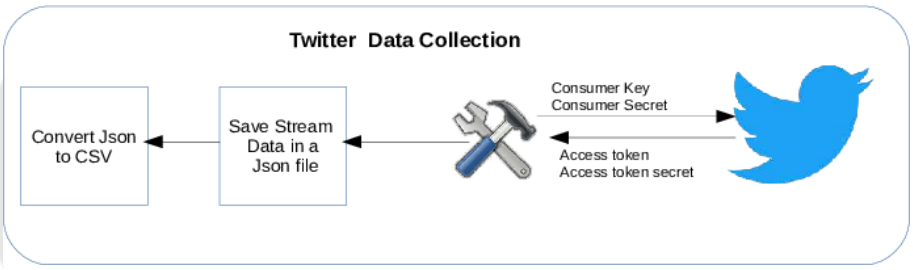
ข้อมูล	ประเภทข้อมูล	แหล่งที่มา	ปีของข้อมูล
ข้อมูลอุบัติเหตุจาก Twitter	ข้อมูลข้อความ	บัญชี JS100Radio และ FM91	ปี 2564 - 2565
ข้อมูลรายชื่อถนนในกรุงเทพมหานคร	ข้อมูลสารสนเทศเชิงเส้น (Vector)	กรมทางหลวงและการทางพิเศษแห่งประเทศไทย	ปี 2560
ข้อมูลอุบัติเหตุย้อนหลัง	ข้อมูลสถิติ	กรมทางหลวง, การทางพิเศษแห่งประเทศไทย และสำนักงานตำรวจแห่งชาติ	ปี 2560 - 2562
ข้อมูลอุบัติเหตุย้อนหลัง	ข้อมูลสถิติ	สำนักงานปลัดกระทรวงคมนาคม	ปี 2564

4. การสืบค้นข้อมูลและการจัดการฐานข้อมูล

การศึกษาครั้งนี้ทำการรวบรวมข้อมูลจากเอกสาร งานวิจัยจากหน่วยงานของรัฐที่เกี่ยวข้อง เพื่อนำระบบการประมวลผลภาษาธรรมชาติและการหาความสัมพันธ์ระหว่างข้อมูลจากข้อมูลสื่อสังคมออนไลน์เพื่อคาดการณ์อุบัติเหตุ โดยทำการแบ่งเป็นหัวข้อการดำเนินการ ดังนี้

4.1 การรวบรวมข้อมูล (Data Collection)

การรวบรวมข้อมูล (Data Collection) รวบรวมข้อมูลจากตำรา เอกสารผลงานวิจัย การขอข้อมูลจากระบบสื่อสังคมออนไลน์ บทความและข้อมูลจากหน่วยงานต่างๆ เพื่อใช้เป็นข้อมูลเบื้องต้นตลอดจนเป็นแนวทางในการศึกษา โดยใช้การนำเข้าข้อมูลผ่านการใช้งาน API ระหว่างผู้ให้บริการและผู้ใช้งาน ดังภาพประกอบ 19 แสดงการเก็บข้อมูลผ่านระบบทวิตเตอร์แบบเรียลไทม์ ทั้งนี้ผู้ใช้งานจำเป็นต้องมีบัญชีสำหรับผู้พัฒนาระบบ (Developer Account) สำหรับการรับส่งข้อมูล



ภาพประกอบ 19 การเก็บข้อมูลผ่านระบบทวิตเตอร์

ที่มา : (S., 2023)

4.2 การจัดการฐานข้อมูลอุบัติเหตุ (Accident Database)

การออกแบบฐานข้อมูลอุบัติเหตุ (Accident Database) โดยการบันทึกการเกิดอุบัติเหตุผ่าน Hashtag ที่มีคำว่าอุบัติเหตุ และจากบัญชีผู้ใช้งาน Twitter ที่มีความเกี่ยวข้องกับการรายงานการจราจรและอุบัติเหตุ เพื่อนำมาเป็นฐานข้อมูลเพื่อนำไปประมวลผลภาษาธรรมชาติและหาความสัมพันธ์ระหว่างข้อความ ดังภาพประกอบ 20

id	username	created_at	tweet	retweet_count	place	location
17	fm91trafficpro	2021-03-06 08:04:51	15.01 น. #อุบัติเหตุ ถนนรามคำแหง ขาออก ช่วงซอยรามคำแหง 83 รถนั่งส่วนบุคคลจอดเสีย ประสาน ค.วิทยุเสรีไทย #FM91 #รายงานจราจร	0	NULL	NULL
103	fm91trafficpro	2021-03-06 09:11:29	15.58 น. #อุบัติเหตุ ถนนกิ่งแก้ว มาจากลาดกระบัง มุ่งหน้า บางนา-ตราด ช่วงซอยกิ่งแก้ว 44 รถชนกัน 3 คัน รถเสียหายมาก อ... https://t.co/faKQ77tsOs	0	NULL	NULL
341	js100radio	2021-03-06 11:47:07	18:46 #อุบัติเหตุ #ถนนรัตนาธิเบศร์ ช่วงสะพานพระนั่งเกล้า > ถ.กาญจนาภิเษก บนสะพานข้ามแยกบางพลู รถชนกัน 5 คัน การจราจรเคลื่อนตัวช้า	0	NULL	Bangkok
430	js100radio	2021-03-06 13:01:58	20:01 #อุบัติเหตุ #ถนนบรมราชชนนี เลยต่างระดับจิมพลี 300 ม. > พุทธมณฑลสาย2 รถเก๋งชนกับรถปิคอัพ การจราจรระยะลอตัว	0	NULL	Bangkok
448	fm91trafficpro	2021-03-06 13:31:26	20.25 น. #อุบัติเหตุ ถนนประเสริฐมนูกิจ ขาเข้า ก่อนถึงแยกได้ด่วนประเสริฐมนูญกิจ ประมาณ 200 เมตร รถเก๋งชนกับรถจักรยาน... https://t.co/4FwhapvzRJ	0	NULL	NULL
661	fm91trafficpro	2021-03-06 16:35:54	23.30 น. ถนนลาดพร้าว ขาออก เลยซอยภาวนา เล็กน้อย รถจักรยานยนต์ชนกัน 2 คัน บาดเจ็บ 2 ราย #FM91 #อุบัติเหตุ #รายงานจราจร	0	NULL	NULL
691	js100radio	2021-03-06 16:53:41	23:53 #ถนนพิษณุโลก ช่วงสนามม้านางเลิ้ง >แยกยมราช ก่อนขึ้นสะพานข้ามแยกยมราชเล็กน้อย รถจักรยานยนต์ชนกับรถเก๋ง ขวางช่องทางกลาง #อุบัติเหตุ	0	NULL	Bangkok
692	fm91trafficpro	2021-03-06 16:54:00	23.48 น. กลางอุโมงค์ลอดแยกเกษตร มุ่งหน้า ถนนประเสริฐมนูกิจ รถจักรยานยนต์พลิกคว่ำ มีคนเจ็บ กีดขวางเลนซ้าย #FM91 #อุบัติเหตุ #รายงานจราจร	0	NULL	NULL
703	fm91trafficpro	2021-03-06 17:02:54	23.54 น. #อุบัติเหตุ สะพานข้ามแยกเจริญราษฎร์ มุ่งหน้า ถนนตก รถนั่งส่วนบุคคล เสียหลักพลิกตะแคง กีดขวางการจราจร ไม่มี... https://t.co/Mol9jte3SY	0	NULL	NULL
712	fm91trafficpro	2021-03-06 17:14:03	00.08 น. กลางห้าแยกลาดพร้าว ถนนวิภาวดี-รังสิต ขาเข้า รถนั่งส่วนบุคคล 2 คันชนกัน ไม่ทราบคนเจ็บ กีดขวางเลนกลาง #FM91 #อุบัติเหตุ #รายงานจราจร	0	NULL	NULL
744	fm91trafficpro	2021-03-06 18:27:43	01.19 น. คืบหน้ากรณี #อุบัติเหตุ สะพานข้ามแยกเจริญราษฎร์ มุ่งหน้า ถนนตก รถนั่งส่วนบุคคล เสียหลักพลิกตะแคง เจ้าหน้า... https://t.co/geU6nk7ZSz	0	NULL	NULL
932	fm91trafficpro	2021-03-06 19:05:07	จยย.ชนเสาแกะกลางถนน เทพารักษ์ กม.24 มีผู้เสียชีวิต 1 คน #FM91 #อุบัติเหตุ #เสียชีวิต #รายงานจราจร https://t.co/V4o6vnXo5V	0	NULL	NULL
1021	fm91trafficpro	2021-03-06 20:50:10	03.48 น. กลางแยกบางยี่เรือ รถจักรยานยนต์พลิกคว่ำ มีคนเจ็บ กีดขวางเลนกลาง #FM91 #อุบัติเหตุ #รายงานจราจร	0	NULL	NULL
1048	fm91trafficpro	2021-03-06 22:33:54	เก๋งชนท้ายรถพ่วง อ.บ้านค่าย จ.ระยอง มีผู้เสียชีวิต 1 คน #FM91 #อุบัติเหตุ #เสียชีวิต #บ้านค่าย #ระยอง https://t.co/Km1bwlnURx	0	NULL	NULL
1086	js100radio	2021-03-07 00:20:12	07:00 #อุบัติเหตุ #ถนนเลียบคลองบางไผ่ จากตัด.กาญจนาภิเษก ตะวันตก >หน้าวัดลาดปลาดุก ที่ไกล้ปากซ.หมู่บ้านบัวทอง16 รถ... https://t.co/Ij6nKmkJ3U	0	NULL	Bangkok
1092	js100radio	2021-03-07 00:35:29	07:25 #อุบัติเหตุ ในซ.สุขุมวิท105 จากแยกตัด.ศรีนครินทร์ >ก่อนถึงรพ.ศิครินทร์ เล็กน้อย ปิดผิวทับจยย.ชนกัน ขวางช่องท... https://t.co/VXqO4vBObQ	0	NULL	Bangkok

ภาพประกอบ 20 ระบบฐานข้อมูลอุบัติเหตุ

5. การประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ

เนื่องจากข้อความในระบบสื่อสังคมออนไลน์ประกอบด้วยข้อความที่เขียนเป็นภาษาไทยและภาษาอังกฤษ ในการประมวลผลภาษาธรรมชาติได้ใช้ไลบราลีสำหรับการประมวลผลภาษาธรรมชาติที่ใช้กับภาษาไทย ไพไทยเอ็นแอลพี (PythaiNLP) โดยมีขั้นตอนการประมวลผลภาษาธรรมชาติ ดังนี้

5.1 การประมวลผลภาษาธรรมชาติจากข้อมูลสื่อสังคมออนไลน์

การประมวลผลภาษาธรรมชาติจากข้อมูลสื่อสังคมออนไลน์ ประกอบด้วย 3 ขั้นตอน ได้แก่

a. การเตรียมข้อมูล (Text Preprocessing phase)

ข้อมูลอุบัติเหตุที่ทำการเก็บรวบรวมได้จะอยู่ในรูปแบบข้อมูลที่ไม่มีโครงสร้างในรูปแบบข้อความ เช่น ตัวอักษรพิเศษ ลิงค์ของเวปไซต์ อิโมจิ และนำมาแปลงข้อมูลให้อยู่ในรูปแบบข้อมูลที่มีโครงสร้างที่แน่นอนโดยการจัดทำฐานข้อมูล โดยมีคำสำคัญในการวิเคราะห์ คือ อุบัติเหตุ ที่ถูกกล่าวถึงในข้อความทั้งหมด ดังภาพประกอบ 21

	tweet	retweet_count	place	\
0	15.01 น. #อุบัติเหตุ ถนนรามคำแหง ขาออก ช่วงซอย...	0	None	
1	15.58 น. #อุบัติเหตุ ถนนกิ่งแก้ว มาจากลาดกระบั...	0	None	
2	18:46 #อุบัติเหตุ #ถนนรัตนาธิเบศร์ ช่วงสะพานพร...	0	None	
3	20:01 #อุบัติเหตุ #ถนนบรมราชชนนี เลยต่างระดับฉ...	0	None	
4	20.25 น. #อุบัติเหตุ ถนนประเสริฐมนูกิจ ขาเข้า ...	0	None	
...	...	...	...	...
9995	02:26 #อุบัติเหตุ #ถนนกาญจนาภิเษก ช่วงถนนบรมรา...	0	None	
9996	03:18 #อุบัติเหตุ #ถนนปู่เจ้าสมิงพราย ช่วงสถาน...	0	None	
9997	03ช26 #อุบัติเหตุ #ถนนลาดพร้าว101 ช่วงตลาดโลตั...	0	None	
9998	03:30 #อุบัติเหตุ #ถนนสาธุประดิษฐ์ ช่วงโรงเรี...	0	None	
9999	03:51 #อุบัติเหตุ #ถนนลาดพร้าว101 ช่วงตลาดโลตั...	0	None	

ภาพประกอบ 21 การเตรียมข้อมูล

b. การกรองข้อมูล (Filtration)

การสกัดคุณลักษณะของข้อความ (Text Extraction) โดยคัดเลือกเฉพาะส่วนที่ต้องการด้วยคำเดี่ยว พยางค์ วลี กลุ่มคำ หรือประโยค และกำจัดคำหยุด (Stopword) โดยตัดคำที่ไม่มีนัยสำคัญต่อเอกสารและไม่ทำให้ใจความของเอกสารเปลี่ยน เช่น คำสรรพนาม คำเชื่อม ชื่อเฉพาะ เช่น ชื่อถนน เป็นต้น (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557) โดยการใช้ไลบราลี (Library) เครื่องมือ PythaiNLP เป็นเครื่องมือในการเรียนรู้ การตัดคำภาษาไทยที่ชื่อว่า newmm engine ตัดคำในขั้นตอนแรก และใช้ thai_stopwords ในการลบคำที่ไม่สำคัญออก ผลที่ได้คือ ชื่อถนนที่เกิดอุบัติเหตุ รายละเอียดเหตุการณ์และความเสียหายที่เกิดขึ้นจากอุบัติเหตุ เพื่อประมวลผลข้อความ (Tokenization) ใช้เครื่องมือ (Engine) การตัดคำด้วยการจับคู่คลังคำศัพท์ (Corpus) และนำ

ข้อมูลที่ไม่เกี่ยวข้องด้วยด้วยนิพจน์ปรกติ (Regular Expression) เช่น การลบข้อมูลที่เป็นตัวเลข ลิงค์ของเวปไซต์ หรือแฮชแท็ก ดังภาพประกอบ 22

```
def replace_unwant_text1(df):
    # regular expression
    df.tweet = df.tweet.replace(to_replace=r'@[a-zA-Z0-9_]{0,15}', value='', regex=True)
    df.tweet = df.tweet.replace(to_replace=r'&gt', value='', regex=True)
    df.tweet = df.tweet.replace(to_replace=r'.\.', value='', regex=True)
    df.tweet = df.tweet.replace(to_replace=r'#อุบัติเหตุ', value='', regex=True)
    df.tweet = df.tweet.replace(to_replace=r'#', value='', regex=True)
    df.tweet = df.tweet.replace(to_replace=r'FM', value='', regex=True)
    df.tweet = df.tweet.replace(to_replace=r'ช่วง', value='', regex=True)
    df.tweet = df.tweet.replace(to_replace=r'RT', value='', regex=True)
    df.tweet = df.tweet.replace(to_replace=r'ขาเข้า', value='', regex=True)
    df.tweet = df.tweet.replace(to_replace=r'ขาออก', value='', regex=True)
    return df
```

ภาพประกอบ 22 การกรองข้อมูลด้วย Regular Expression

c. การทำกลุ่มข้อมูล (Word Cloud) และความถี่ข้อมูล (Frequency)

การวิเคราะห์ข้อมูลจากจำนวนความถี่ของคำที่เกิดขึ้นซ้ำๆ จากการรวบรวมข้อมูลจากสื่อสังคมออนไลน์ให้อยู่ในรูปแบบข้อมูลก้อนเมฆ (Word Cloud) โดยใช้เครื่องมือที่ช่วยในการแสดงภาพความถี่ของข้อความ และแผนภูมิแสดงความถี่ของข้อมูล (Frequency of word) ดังภาพที่ 23 เมื่อได้กลุ่มข้อความก้อนเมฆที่แสดงความถี่ของถนนที่เกิดอุบัติเหตุสูงสุดจากนั้นนำรายชื่อถนนที่ได้ทำการเชื่อมโยงกับคำที่มีความถี่สูงสุดบนถนนนั้นๆ

```
def create_wordcloud(data):
    data = replace_unwant_text(data)
    data = all_row_into_one(data)
    #ทำการ tokenization จากข้อมูล all_tweet เป็น arrey ที่ตัดคำของ all_tweet
    d = sent_tokenize(data, engine="whitespace+newline")
    #แปลง arrey เป็น dataframe
    #text = pd.DataFrame(data=d)

    #ทำ
    wordcloud = WordCloud(width=1000, height=600,
                          font_path='/home/jovyan/Sitthipat/THSarabunNew.ttf',
                          background_color="#F0FFFF",
                          regexp=r"[\u0E00-\u0E7Fa-zA-Z']+").generate(' '.join(d))
    #wordcloud = WordCloud().generate(tweet)
    #กำหนดขนาดกรอบ
    plt.figure( figsize=(20,10))
    #เอา wordcloud มาใส่ในกรอบ plt
    plt.imshow(wordcloud, interpolation='bilinear')
    #ซ่อนแกนกราฟ
    plt.axis("off")
    #แสดงภาพ
    plt.show()
```

ภาพประกอบ 23 รูปแบบข้อมูลก้อนเมฆ (Word Cloud)

6. การหาความสัมพันธ์ระหว่างข้อมูลจากการประมวลผลภาษาธรรมชาติ

การหาความสัมพันธ์ของเหมืองดัชนีถ้อยคำ (Association rule mining) เป็นกระบวนการค้นหาความสัมพันธ์ระหว่างเหมืองดัชนีถ้อยคำ (Term frequency) ในกลุ่มข้อมูลข้อความ ความสัมพันธ์เหล่านี้มักถูกแสดงในรูปของกฎ (rule) ซึ่งมีรูปแบบดังนี้

เหมืองดัชนีถ้อยคำหลักเป็นคำว่า “ชื่อถนนที่เกิดอุบัติเหตุสูงสุด” และเหมืองดัชนีถ้อยคำรองเป็นคำว่า “อุบัติเหตุ” ซึ่งจะนำมาหาความสัมพันธ์ของการพบคำเหล่านี้ด้วยกัน

7. กระบวนการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์

การแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์สำหรับ python สามารถเลือกใช้ได้หลากหลาย ตัวอย่างหนึ่งที่สะดวกและง่าย คือการใช้ผ่าน Geopandas ซึ่งมีฟังชั่น geocode() ที่สามารถแปลงที่อยู่ไปเป็นวัตถุรูปทรงเรขาคณิต (geometry object) ข้อมูลประเภทจุด (Point Data) ที่มีพิกัดแผนที่อ้างอิงได้ทันที โดยมีกระบวนการดังนี้

7.1 การนำเข้าโมดูล เครื่องมือที่สำคัญในการทำ Geocoding

ด้วยภาษา Python ต้องมีการทำให้ตัวเครื่องคอมพิวเตอร์ที่ประมวลผล รับรู้โมดูล เครื่องมือที่จำเป็นในการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์ด้วยเครื่องมือ pandas และ geopandas ในทำงานด้านการประมวลผลข้อมูลเชิงพื้นที่ (Geospatial data) ทำหน้าที่จัดการข้อมูลประเภทเรขาคณิต (geometry data type) โดยจะเชื่อมกับข้อมูลเชิงบรรยายประกอบต่างๆ ข้อดีของ GeoPandas สามารถประมวลผลข้อมูลเรขาคณิตได้ทำให้มีความสามารถรองรับการทำ Geo processing เช่นการสร้างแนวระยะห่าง (buffer) การหาพื้นที่ซ้อนทับ (intersection) การเชื่อมต่อข้อมูล (Merge) การหาพื้นที่ซ้อน (Union) รวมถึงการคำนวณบนข้อมูลประเภทเรขาคณิต การหาระยะ หาพื้นที่ การจัดการเรื่องระบบอ้างอิงพิกัด (Coordinate reference system) โดยข้อมูลจัดการในรูปแบบของข้อมูลที่รองรับการวิเคราะห์และประมวลผลข้อมูล ด้วย PANDAS บนภาษา Python ได้ ดังภาพประกอบ 24

```
import pandas as pd
import geopandas as gpd
from shapely.geometry import Point
from geopandas.tools import geocode
```

ภาพประกอบ 24 การเลือกโมดูลในการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์

7.2 การเตรียมข้อมูลที่ทำการบันทึกจาก Twitter

การเตรียมข้อมูลที่ทำการบันทึกจาก Twitter ที่อยู่ในรูปแบบข้อความ เพื่อใช้แปลงเป็นพิกัดผ่านฟังชั่น geocode() โดยมีการนำเข้าข้อมูลในรูปแบบข้อความ เพื่อเตรียมนำไปแปลงเป็นระบบพิกัดทางภูมิศาสตร์ที่ได้มาจากการแปลงข้อความที่เป็นชื่อถนน ดังภาพประกอบ 25

```
data = {'Site 1':'ถนนสารสิน, Witthayu, แขวงลุมพินี, เขตปทุมวัน, กรุงเทพมหานคร, 10330, ประเทศไทย',
     'Site 2':'Kamou Tokzung(豬腳飯), Soi Charoen Wiang, Sathon, แขวงสีลม, เขตบางรัก, กรุงเทพมหานคร, 10500, ประเทศไทย',
      'Site 3': 'โรงเรียนสาธิตจุฬาลงกรณ์มหาวิทยาลัย ฝ่ายมัธยม, ซอยจุฬาลงกรณ์ 42, Sam Yan, แขวงวังใหม่, เขตปทุมวัน, กรุงเทพมหานคร, 10330, ประเทศไทย',
      'Site 4': 'วัดหัวลำโพง, ซอยหน้าวัดหัวลำโพง, แขวงสุริยวงศ์, เขตบางรัก, กรุงเทพมหานคร, 10500, ประเทศไทย',
      'Site 5': 'โรงพยาบาลจุฬาลงกรณ์, ถนนอังรีดูนังต์, Siam, แขวงปทุมวัน, เขตปทุมวัน, กรุงเทพมหานคร, 10330, ประเทศไทย ',
      'Site 6': 'Suphachalasai Stadium, ซอยจุฬาลงกรณ์ 5, แขวงวังใหม่, เขตปทุมวัน, กรุงเทพมหานคร, 10330, ประเทศไทย'
}

df = pd.DataFrame.from_dict(data, orient='index')
df.head()
```

ภาพประกอบ 25 การนำข้อมูลมาทำการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์

7.3 การเรียกใช้เครื่องมือโมดูล geocode ใน Python

การเรียกใช้เครื่องมือ geocode เพื่อแปลงข้อมูลที่อยู่ (address) ให้เป็นข้อมูลรูปทรงเรขาคณิตหรือข้อมูลจุด (Geometry Point Object) โดยระบบจะส่งค่ากลับเป็นพิกัดจุดแบบ 2 มิติ (X,Y Point) และนำมาทำการแปลงจากพิกัดจุด เป็นเส้นถนนโดยเชื่อมโยงจากชื่อถนน (Line) ในการแปลงให้เป็นระบบพิกัดทางภูมิศาสตร์ โดยเลือกใช้เครื่องมือ nominatim ซึ่งเป็นผู้ให้บริการภายใต้ Openstreetmap ดังภาพประกอบ 26

```
geo = geocode(df[0], provider='nominatim')
geo.head()
```

ภาพประกอบ 26 การเรียกใช้ฟังก์ชันการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์

7.4 ผลลัพธ์ที่ได้จากการประมวลผล Geocoding

ในการแปลงข้อความเป็นพิกัดทางภูมิศาสตร์ จะได้ค่าพิกัดทางภูมิศาสตร์ของสถานที่ เช่น ถนนสารสิน จะได้การแปลงข้อมูลเป็นพิกัดทางภูมิศาสตร์ ดังภาพประกอบ 27

	address	geometry
Site 1	ถนนสารสิน, วิทยุ, แขวงลุมพินี, เขตปทุมวัน, กรุ...	POINT (100.5421532 13.7341508)
Site 2	Kamou Tokzung(豬腳飯), Soi Charoen Wiang, สาทร, แ...	POINT (100.5164317 13.719473)
Site 3	โรงเรียนสาธิตจุฬาลงกรณ์มหาวิทยาลัย ฝ่ายมัธยม, ...	POINT (100.527247839107 13.73607025)
Site 4	วัดหัวลำโพง, ซอยหน้าวัดหัวลำโพง, แขวงสุริยวงศ์...	POINT (100.529371082842 13.7316989)
Site 5	โรงพยาบาลจุฬาลงกรณ์, ถนนอังรีดูนังต์, สยาม, แข...	POINT (100.535638271553 13.73273225)

ภาพประกอบ 27 ผลลัพธ์จากการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์

8. การจัดทำแผนที่ความเสี่ยงจากการเกิดอุบัติเหตุ ตามช่วงเวลา

การสร้างแผนที่การเกิดอุบัติเหตุในแต่ละช่วงเวลา จากการรวบรวมข้อมูลสภาพการจราจรย้อนหลัง พบว่าช่วงเช้า จะเริ่มมีการจราจรหนาแน่นตั้งแต่ช่วงเช้าเวลา 06.00 น. ถึง 9.00 น. ช่วงกลางวัน เวลา 11.00 น. ถึง 14.00 น. และช่วงเย็นตั้งแต่ 16.00 น. ถึง 20.00 น.

การประเมินความเสี่ยงของการเกิดอุบัติเหตุ (Evaluation of accident) ขั้นตอนนี้เป็นการนำพิกัดทางภูมิศาสตร์ (Latitude, Longitude) จากการทำกระบวนการแปลงข้อความเป็นพิกัดทางภูมิศาสตร์ (geocoding) มาทำการประมาณเส้น (Line) ของถนนที่มีความถี่ในการเกิดอุบัติเหตุลงบนแผนที่ เพื่อหาพื้นที่เสี่ยงของการเกิดอุบัติเหตุ การวิเคราะห์ข้อมูลเพื่อหาบริเวณที่เกิดอุบัติเหตุจำนวนมากหรือบริเวณที่เสี่ยงต่อการเกิดอุบัติเหตุ

9. การตรวจสอบความสอดคล้องกับข้อมูลสถิติการเกิดอุบัติเหตุ

การตรวจสอบความสอดคล้องของข้อมูลที่บันทึกจากสื่อสังคมออนไลน์ โดยการเปรียบเทียบข้อมูลจากศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงคมนาคม ที่มีรายงานการเกิดอุบัติเหตุในช่วง ปี 2564 ที่อยู่ในช่วงเวลาเดียวกันกับที่บันทึกข้อมูลจากสื่อสังคมออนไลน์ เพื่อตรวจสอบความถูกต้องและความน่าเชื่อถือของชุดข้อมูลที่ได้ทำการวิเคราะห์จากสื่อสังคมออนไลน์

บทที่ 4
ผลการดำเนินงานวิจัย

การวิจัยครั้งนี้เป็นการศึกษากระบวนการประมวลผลภาษาธรรมชาติ จากข้อมูลสื่อสังคมออนไลน์ เพื่อคาดการณ์อุบัติเหตุเชิงพื้นที่ โดยการศึกษาความสัมพันธ์ระหว่างข้อมูล ผลการวิจัยสามารถดำเนินการวิเคราะห์ข้อมูล 4 ขั้นตอน ดังนี้

1. ฐานข้อมูลการเกิดอุบัติเหตุ (Accident Database)

2. การประมวลภาษาธรรมชาติและการวิเคราะห์เหมืองดัชนีถ้อยคำของข้อมูลการเกิดอุบัติเหตุจากสื่อสังคมออนไลน์ (Natural Language Processing of Accident in social media)

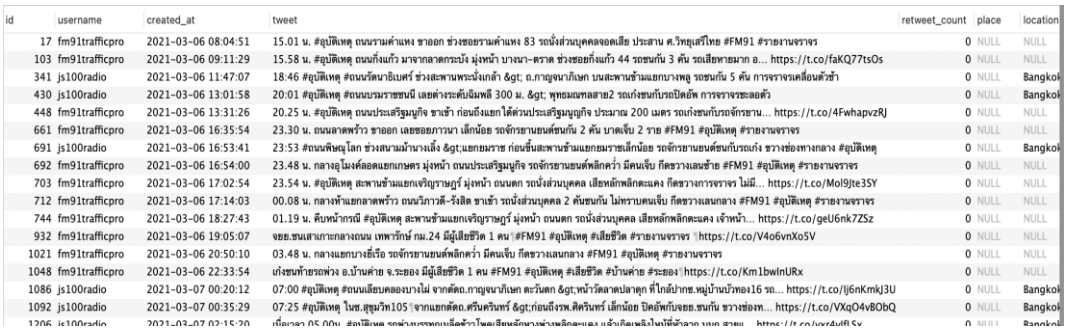
3. การวิเคราะห์เหมืองข้อมูลกับสถิติการเกิดอุบัติเหตุ

4. การวิเคราะห์เหมืองข้อมูลและการหาความสัมพันธ์ระหว่างข้อมูลการเกิดอุบัติเหตุด้วยอัลกอริทึมอพริโอริ (Association Rule Mining by Apriori Algorithm)

5 การแปลงข้อความการเกิดอุบัติเหตุด้วยเป็นระบบพิกัดทางภูมิศาสตร์ (Geocoding of Accident Location)

1. ฐานข้อมูลการเกิดอุบัติเหตุ (Accident Database)

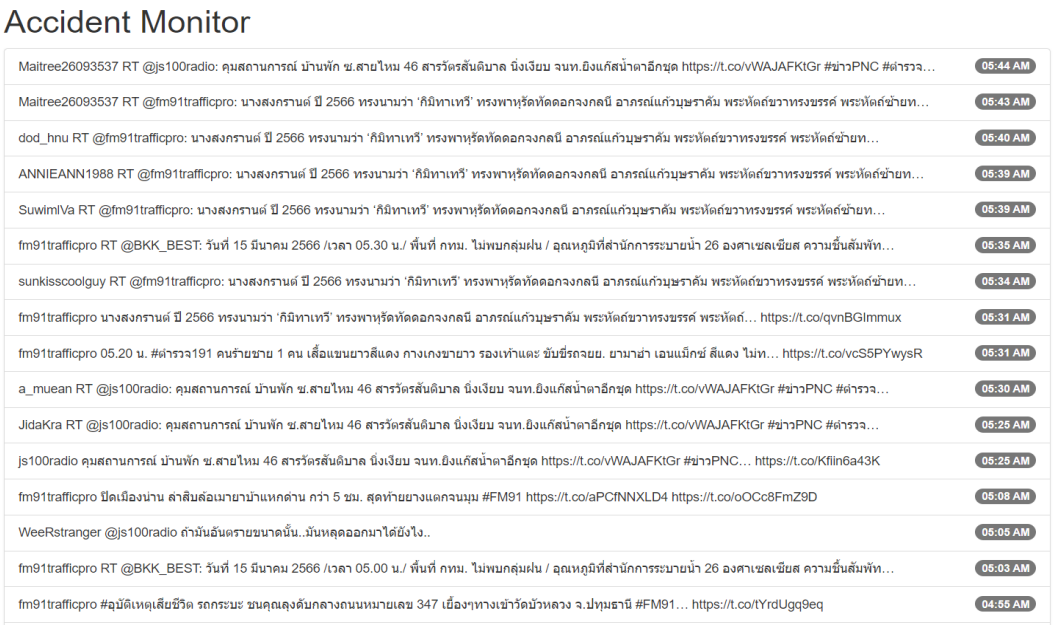
การสร้างคลังฐานข้อมูลเชิงพื้นที่ของอุบัติเหตุจากสื่อสังคมออนไลน์ (Database) งานวิจัยนี้รวบรวมข้อมูลโดยการสืบค้นข้อมูลจากสื่อสังคมออนไลน์ Twitter ผ่านการเรียกใช้งาน API และสร้างคลังฐานข้อมูลในรูปแบบ MySQL เพื่อจัดเก็บและดึงข้อมูลทั้งหมดรวมถึงเนื้อหาโพสต์ โปรไฟล์ผู้ใช้ และประเภทโพสต์ที่กำหนดเอง โดยกำหนดเงื่อนไขในการเก็บข้อมูล โดยข้อมูลที่ได้ส่วนใหญ่จะอยู่ในพื้นที่บริเวณกรุงเทพมหานคร ช่วงปี พ.ศ.2563 จากบัญชีที่มีการรายงานสภาพอุบัติเหตุและจราจร 2 บัญชี คือ @JS100Radio และ @FM91radiopro โดยเลือกข้อมูลผ่านแฮชแท็ก (Hash tag) ที่ใช้คำว่า “อุบัติเหตุ” เป็นข้อความที่มีข้อมูลเกี่ยวกับเหตุการณ์ที่สามารถนำไปประมวลผลภาษาธรรชาติต่อไป ข้อมูลที่ได้อยู่ในรูปแบบเอกสารที่เป็นตัวอักษรและรองรับการเข้ารหัส (Unicode) ได้ทุกภาษา ข้อมูลประกอบไปด้วย ชื่อผู้ใช้งาน (Username) วันและเวลาที่ข้อมูลอัปโหลด (Date and Time) ดัชนีถ้อยคำ (Tags) เพื่อนำไปเป็นข้อมูลในการวิเคราะห์และหาความสัมพันธ์ของข้อความ ข้อมูลถูกเก็บในฐานข้อมูล MySQL ดังภาพประกอบ 28



id	username	created_at	tweet	retweet_count	place	location
17	fm91trafficpro	2021-03-06 08:04:51	15.01 น. #อุบัติเหตุ ถนนรามคำแหง ขาออก ช่วงซอยรามคำแหง 83 รถนั่งส่วนบุคคลจอดเสีย ประสาน ศ.วิทยุเสรีไทย #FM91 #รายงานจราจร	0	NULL	NULL
103	fm91trafficpro	2021-03-06 09:11:29	15.58 น. #อุบัติเหตุ ถนนกิ่งแก้ว มาจากลาดกระบัง มุ่งหน้า บางนา-ตราด ช่วงซอยกิ่งแก้ว 44 รถชนกัน 3 คัน รถเสียหายมาก อ... https://t.co/faKQ77tsOs	0	NULL	NULL
341	js100radio	2021-03-06 11:47:07	18:46 #อุบัติเหตุ #ถนนรัตนาธิเบศร์ ช่วงสะพานพระนั่งเกล้า > อ.กาญจนาภิเษก บนสะพานข้ามแยกบางพลู รถชนกัน 5 คัน การจราจรเคลื่อนตัวช้า	0	NULL	Bangkok
430	js100radio	2021-03-06 13:01:58	20:01 #อุบัติเหตุ #ถนนบรมราชชนนี เลยต่างระดับจิมพลี 300 ม. > พุทธมณฑลสาย2 รถเก๋งชนกับรถปิคอัพ การจราจรปะลอตัว	0	NULL	Bangkok
448	fm91trafficpro	2021-03-06 13:31:26	20.25 น. #อุบัติเหตุ ถนนประเสริฐมนูกิจ ขาเข้า ก่อนถึงแยก ได้ส่วนประเสริฐมนูกิจ ประมาณ 200 เมตร รถเก๋งชนกับรถจักรยาน... https://t.co/4FwhapvzRJ	0	NULL	NULL
661	fm91trafficpro	2021-03-06 16:35:54	23.30 น. ถนนลาดพร้าว ขาออก เลยซอยภาวนา เล็กน้อย รถจักรยานยนต์ชนกัน 2 คัน บาดเจ็บ 2 ราย #FM91 #อุบัติเหตุ #รายงานจราจร	0	NULL	NULL
691	js100radio	2021-03-06 16:53:41	23:53 #ถนนพิษณุโลก ช่วงสนามม้านางเลิ้ง >แยกยมราช ก่อนขึ้นสะพานข้ามแยกยมราชเล็กน้อย รถจักรยานยนต์ชนกับรถเก๋ง ขวางช่องทางกลาง #อุบัติเหตุ	0	NULL	Bangkok
692	fm91trafficpro	2021-03-06 16:54:00	23.48 น. กลางอุโมงค์ลอดแยกเกษตร มุ่งหน้า ถนนประเสริฐมนูกิจ รถจักรยานยนต์หลักคว่ำ มีคนเจ็บ กีดขวางเลนซ้าย #FM91 #อุบัติเหตุ #รายงานจราจร	0	NULL	NULL
703	fm91trafficpro	2021-03-06 17:02:54	23.54 น. #อุบัติเหตุ สะพานข้ามแยกเจริญราษฎร์ มุ่งหน้า ถนนตก รถนั่งส่วนบุคคล เสียหลักพลิกตะแคง กีดขวางการจราจร ไม่มี... https://t.co/Mol9jte3SY	0	NULL	NULL
712	fm91trafficpro	2021-03-06 17:14:03	00.08 น. กลางห้าแยกลาดพร้าว ถนนวิภาวดี-รังสิต ขาเข้า รถนั่งส่วนบุคคล 2 คันชนกัน ไม่ทราบคนเจ็บ กีดขวางเลนกลาง #FM91 #อุบัติเหตุ #รายงานจราจร	0	NULL	NULL
744	fm91trafficpro	2021-03-06 18:27:43	01.19 น. คืบหน้ากรณี #อุบัติเหตุ สะพานข้ามแยกเจริญราษฎร์ มุ่งหน้า ถนนตก รถนั่งส่วนบุคคล เสียหลักพลิกตะแคง เจ้าหน้า... https://t.co/geU6nk7ZSz	0	NULL	NULL
932	fm91trafficpro	2021-03-06 19:05:07	จยย.ชนเสาไกรกลางถนน เทพารักษ์ กม.24 มีผู้เสียชีวิต 1 คน #FM91 #อุบัติเหตุ #เสียชีวิต #รายงานจราจร https://t.co/V4o6vnXo5V	0	NULL	NULL
1021	fm91trafficpro	2021-03-06 20:50:10	03.48 น. กลางแยกบางยี่เรือ รถจักรยานยนต์หลักคว่ำ มีคนเจ็บ กีดขวางเลนกลาง #FM91 #อุบัติเหตุ #รายงานจราจร	0	NULL	NULL
1048	fm91trafficpro	2021-03-06 22:33:54	เก่งชนท้ายรถพ่วง อ.บ้านค่าย จ.ระยอง มีผู้เสียชีวิต 1 คน #FM91 #อุบัติเหตุ #เสียชีวิต #บ้านค่าย #ระยอง https://t.co/Km1bwlnURx	0	NULL	NULL
1086	js100radio	2021-03-07 00:20:12	07:00 #อุบัติเหตุ ถนนเลียบคลองบางไผ่ จากตัด.กาญจนาภิเษก ตะวันตก >หน้าวัดลาดปลาดุก ที่ใกล้ปากซ.หมู่บ้านบัวทอง16 รถ... https://t.co/Ij6nKmkJ3U	0	NULL	Bangkok
1092	js100radio	2021-03-07 00:35:29	07:25 #อุบัติเหตุ ในซ.สุขุมวิท105 จากแยกตัด.ศรีนครินทร์ >ก่อนถึงรพ.ศิครินทร์ เล็กน้อย ปิดอัพกับจยย.ชนกัน ขวางช่องท... https://t.co/VXqO4vBObQ	0	NULL	Bangkok
1206	js100radio	2021-03-07 02:15:20	เมื่อเวลา 05.00น. #อุบัติเหตุ รถพ่วงบรรทุกเมล็ดข้าวโพดเสียหลักขวางช่วงพลิกตะแคง แล้วเกิดพลิกไหม่ที่ด้านขาล... https://t.co/yxr4vlfLSx	0	NULL	Bangkok

ภาพประกอบ 28 ระบบฐานข้อมูล

การพัฒนาโปรแกรมเพื่อทำการสืบค้นข้อมูลดัชนีถ้อยคำและจัดทำคลังฐานข้อมูลจากสื่อสังคมออนไลน์ ถูกพัฒนาโดยโปรแกรมภาษา Python ด้วยเครื่องมือ Tweepy ในการบันทึกข้อมูลจากสื่อสังคมออนไลน์ Twitter โดยเก็บลงในฐานข้อมูล Mysql เพื่อนำไปเป็นฐานข้อมูลสำหรับการประมวลผลภาษาธรรมชาติ โดยจัดเก็บข้อมูลจากบัญชีผู้ใช้งานทั้งหมดที่มีการกล่าวถึง จส.100 และ สวพ.91 ด้วย ดังภาพประกอบ 29 เป็นการจัดทำรูปแบบการแสดงผลผ่านเว็ปบราวเซอร์ที่ได้จัดทำขึ้นเพื่อตรวจสอบการทำงานของระบบดึงข้อมูลแบบเรียลไทม์ โดยให้แสดงผลชื่อบัญชีผู้ทวิต ข้อความที่ทวีต และเวลาที่ทวีต



ภาพประกอบ 29 การแสดงผลการบันทึกข้อมูลที่ได้จาก Twitter และแสดงผลบนบราวเซอร์

การพัฒนาโปรแกรมเพื่อทำการสืบค้นข้อมูลดัชนีถ้อยคำและจัดทำคลังฐานข้อมูลจาก Twitter ระหว่างปี 2564 – 2565 โดยมีกลุ่มตัวอย่างในการดึงข้อมูลและจัดเก็บเป็นฐานข้อมูลดังภาพประกอบ 32 มีชุดข้อมูลที่ผู้ใช้งานทำการทวีตทั้งสิ้น จำนวน 94,555 ชุดข้อมูล ซึ่งเป็นข้อมูล

ทั้งหมดที่ได้ทำการบันทึกโดยยังไม่ถูกกรองข้อมูลซึ่งจะมีข้อมูลที่ซ้ำกันจากการรีทวีต (Retweet) ที่เป็นข้อมูลจากบัญชีผู้ใช้งานอื่น และข้อมูลที่ไม่เกี่ยวข้องกับอุบัติเหตุ ข้อมูลที่ในคลังฐานข้อมูลที่รวบรวมได้ ประกอบไปด้วย หมายเลขข้อมูล (ID) ชื่อผู้ใช้งาน (Username) วันเวลาที่ทวีตข้อความ (Create_date) ข้อความที่ถูกทวีต (Tweet) พื้นที่ที่ทวิตข้อความ (Location) เพื่อสร้างฐานข้อมูลในรูปแบบ MySQL เป็นข้อมูลแบบมีโครงสร้าง และนำข้อมูลไปประมวลผลภาษาธรรมชาติด้วยการประยุกต์ใช้ภาษา Python และการเชื่อมต่อข้อมูลจากฐานข้อมูล เมื่อทำการกรองข้อมูลจากบัญชีผู้ใช้งานที่ต้องการและเกี่ยวข้องกับอุบัติเหตุแล้วจะได้ 13,095 ชุดข้อความ ดังภาพประกอบ 30-31

```
from sqlalchemy import create_engine
import pandas as pd
import numpy as np
from pythainlp import sent_tokenize
from pythainlp import word_tokenize
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
import matplotlib.pyplot as plt
import re
from collections import Counter
import matplotlib.font_manager as font_manager


#Connect Server
db_connection_str = 'mysql+pymysql://[redacted]/twitter_accident'
db_connection = create_engine(db_connection_str, encoding="utf16")

#Query ข้อมูลไว้ที่ตัวแปร
tw = pd.read_sql("select * from twitter_accidents_filtered limit 10000", con=db_connection)
rama2 = pd.read_sql("select * from hash_rama2_road limit 10000", con=db_connection)
twyear = pd.read_sql("select * from twitter_year limit 1000", con=db_connection)

morning = pd.read_sql("select * from twitter_Accident_Morning limit 10000", con=db_connection)
noon = pd.read_sql("select * from twitter_Accident_Noon limit 10000", con=db_connection)
evening = pd.read_sql("select * from twitter_Accident_Evening limit 10000", con=db_connection)
```

```
<bound method NDFrame.head of            id        username          created_at  \
0          17  fm91trafficpro 2021-03-06 08:04:51
1         103  fm91trafficpro 2021-03-06 09:11:29
2         341      js100radio 2021-03-06 11:47:07
3         430      js100radio 2021-03-06 13:01:58
4         448  fm91trafficpro 2021-03-06 13:31:26
...       ...             ...                 ...
9995  1207327      js100radio 2022-01-29 02:26:25
9996  1207336      js100radio 2022-01-29 03:19:15
9997  1207337      js100radio 2022-01-29 03:27:01
9998  1207340      js100radio 2022-01-29 03:30:46
9999  1207343      js100radio 2022-01-29 03:52:06

                                                tweet  retweet_count place  \
0     15.01 น. #อุบัติเหตุ ถนนรามคำแหง ขาออก ช่วงซอย...              0  None
1     15.58 น. #อุบัติเหตุ ถนนกิ่งแก้ว มาจากลาดกระบั...              0  None
2     18:46 #อุบัติเหตุ #ถนนรัตนาธิเบศร์ ช่วงสะพานพร...              0  None
3     20:01 #อุบัติเหตุ #ถนนบรมราชชนนี เลยต่างระดับฉ...              0  None
4     20.25 น. #อุบัติเหตุ ถนนประเสริฐมนูกิจ ขาเข้า ...              0  None
...                                                 ...            ...   ...
9995  02:26 #อุบัติเหตุ #ถนนกาญจนาภิเษก ช่วงถนนบรมรา...              0  None
9996  03:18 #อุบัติเหตุ #ถนนปู่เจ้าสมิงพราย ช่วงสถาน...              0  None
9997  03ช26 #อุบัติเหตุ #ถนนลาดพร้าว101 ช่วงตลาดโลต้...              0  None
9998  03:30 #อุบัติเหตุ #ถนนสาธุประดิษฐ์ ช่วงโรงเรี...              0  None
9999  03:51 #อุบัติเหตุ #ถนนลาดพร้าว101 ช่วงตลาดโลต้...              0  None
```

ภาพประกอบ 30 การเชื่อมต่อข้อมูลจาก MySQL เพื่อนำมาประมวลผลภาษาธรรมชาติ

id	username	created_at	tweet	retweet_count	place	location	timestamp
1383	fm91trafficpro	2021-03-07 03:29:54	บริษัทกลางฯ เผยข้อมูลรับแจ้งอุบัติเหตุทางถนน ในรอบ 24 ชั่วโมง ประจำวันเสาร์ ที่ 6 มี...	0	NULL	NULL	2021-03-07 10:29:59
1442	js100radio	2021-03-07 04:03:09	11.01 #อุบัติเหตุ #ถนนราชพฤกษ์ ช่วง เ...น้อย >แยกถ.บรมราชชนนี ร...	0	NULL	Bangkok	2021-03-07 11:03:14
1444	js100radio	2021-03-07 04:04:35	11.01 #อุบัติเหตุ #ถนนราชพฤกษ์ จากแยกบางขุนกอง >ช่วงใกล้แยกสวนผัก รถเก๋งเสี...	0	NULL	Bangkok	2021-03-07 11:04:40
1508	js100radio	2021-03-07 04:39:30	11:32 #อุบัติเหตุ #ถนน346 ช่วงตัดทล.9 >ถ.340 เลยแยกนพวงศ์ เล็กน้อย เก๋งกับปิค...	0	NULL	Bangkok	2021-03-07 11:39:35
1531	js100radio	2021-03-07 04:58:11	11:57 #อุบัติเหตุ #ถนนกาญจนาภิเษก ช่วงคลองถนน >ช.วัดลาดปลาดุก ที่ใกล้MRTค...	0	NULL	Bangkok	2021-03-07 11:58:16
1535	fm91trafficpro	2021-03-07 05:00:40	12.00 น. ถนนร่มเกล้า ขาออก ปากซอยร่มเกล้า 21/6 ก่อนหน้านี้มีอุบัติเหตุรถบรรทุกเสียห...	0	NULL	NULL	2021-03-07 12:00:45
1583	fm91trafficpro	2021-03-07 05:36:32	RT @tossapon2535: @fm91trafficpro อุบัติเหตุสามแยก ทางเข้าวัดบางปรง ถนนบ...	0	NULL	NULL	2021-03-07 12:36:37
1593	js100radio	2021-03-07 05:46:39	12:45 #อุบัติเหตุ #ถนนราชพฤกษ์ ช่วงเลยจากอุโมงค์ลอดวงเวียนบางขุนกอง ประมาณ 20...	0	NULL	Bangkok	2021-03-07 12:46:44
1599	js100radio	2021-03-07 05:50:46	12:49 #อุบัติเหตุ #ถนนวิภาวดีรังสิต ช่วงเชิงทางลงสะพานข้ามแยกลาดพร้าว >แยกสุทธิ...	0	NULL	Bangkok	2021-03-07 12:50:51
1642	js100radio	2021-03-07 06:36:34	13:32 #อุบัติเหตุ #ถนนกาญจนาภิเษก ช่วงต่างระดับเอกชัย >ต่างระดับบางโคลัด ที่ตร...	0	NULL	Bangkok	2021-03-07 13:36:39
1657	js100radio	2021-03-07 06:47:14	13.44 #อุบัติเหตุ รถบรรทุกข้าวเปลือกเสียหลักตกข้างทางพลิกตะแคง #ทล12 ช่วงโค้งโมอ...	0	NULL	Bangkok	2021-03-07 13:47:20
1670	js100radio	2021-03-07 06:56:37	RT @kunpapaa: @js100radio มีอุบัติเหตุรถโตโยต้า Prius สีบอร์นเงิน ทะเบียน ญฮ (จ้า...	0	NULL	Bangkok	2021-03-07 13:56:42
2979	fm91trafficpro	2021-03-08 03:41:34	บริษัทกลางฯ เผยข้อมูลรับแจ้งอุบัติเหตุทางถนน ในรอบ 24 ชั่วโมง ประจำวันอาทิตย์ ที่ 7...	0	NULL	NULL	2021-03-08 10:41:39
2983	js100radio	2021-03-08 03:46:29	10:46 #อุบัติเหตุ #ถนนรัชดาภิเษก ช่วง แยกรัชวิภา >ที่ในอุโมงค์รัชโยธิน รถบิคอัพกับ...	0	NULL	Bangkok	2021-03-08 10:46:34
3174	js100radio	2021-03-08 05:34:50	12:34 #อุบัติเหตุ #ถนนลาดพร้าว ช่วง ซอยลาดพร้าว101 >แยกแฮปปี้แลนด์ ที่ปากซ.ล...	0	NULL	Bangkok	2021-03-08 12:34:55
3208	fm91trafficpro	2021-03-08 05:50:52	12.50 น. ถนนมอเตอร์เวย์ (ทล.7) ขาออก มุ่งหน้ามาบตาพุด ช่วงกม.8+800 มีอุบัติเหตุ ก็...	0	NULL	NULL	2021-03-08 12:50:58
3217	fm91trafficpro	2021-03-08 05:52:35	12.52 น. ถนนลาดพร้าว ขาออก ปากซอยลาดพร้าว 101/2 อุบัติเหตุ รถจักรยานยนต์ ชนกั...	0	NULL	NULL	2021-03-08 12:52:40
3218	fm91trafficpro	2021-03-08 05:53:34	12.52 น. ถนนลาดพร้าว ขาออก ปากซอยลาดพร้าว 101/2 อุบัติเหตุ รถจักรยานยนต์ ชนกั...	0	NULL	NULL	2021-03-08 12:53:39
4964	fm91trafficpro	2021-03-09 03:13:22	10.11 น. อุบัติเหตุ 4 คัน อุโมงค์ลอดแยกรัชโยธิน ถนนรัชดาภิเษก ขาเข้า เคลื่อนย้ายไม่ได้...	0	NULL	NULL	2021-03-09 10:13:27
4971	js100radio	2021-03-09 03:22:28	10:22 #อุบัติเหตุ #ถนนเทพรัตน ช่วง สนามบินสุวรรณภูมิ >ม.หัวเฉียวฯ ที่บริเวณกม.17...	0	NULL	Bangkok	2021-03-09 10:22:34
4975	js100radio	2021-03-09 03:26:39	10:26 #อุบัติเหตุ #ถนนอ่อนนุช ช่วง แยกอ่อนนุช44 >แยกอ่อนนุช17 ที่บริเวณช.อ่อนนุ...	0	NULL	Bangkok	2021-03-09 10:26:44

ภาพประกอบ 31 ตัวอย่างคลังฐานข้อมูลในรูปแบบ MySQL จากข้อความบน Twitter ในปี 2564-2565

2. การประมวลภาษาธรรมชาติและการวิเคราะห์เหมืองดัชนีถ้อยคำของข้อมูลการเกิดอุบัติเหตุ (Natural Language Processing of Accident in social media)

2.1 ผลลัพธ์การประมวลผลภาษาธรรมชาติ

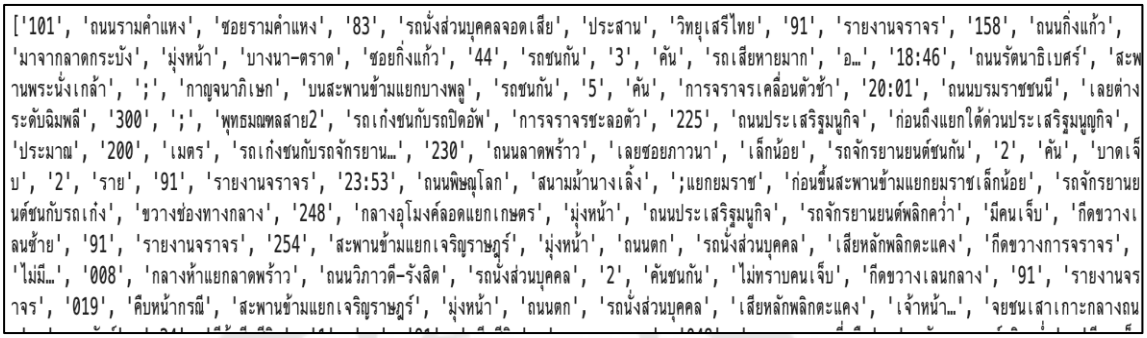
การกรองข้อมูลจากข้อมูลสื่อสังคมออนไลน์เพื่อนำไปประมวลผลภาษาธรรมชาติด้วยโปรแกรมภาษา Python ด้วยเครื่องมือโมดูล PythaiNLP ด้วยเครื่องมือ newmm engine เป็นการตัดคำในขั้นตอนแรก และใช้ thai_stopwords ในการลบคำที่ไม่สำคัญออก ผลที่ได้คือ ชื่อถนนที่เกิดอุบัติเหตุ รายละเอียดเหตุการณ์และความเสียหายที่เกิดขึ้นจากอุบัติเหตุ ดังภาพประกอบ 32

```
['13:32', 'ถนนกาญจนาภิเษก', 'ต่างระดับเอกชัย', ';ต่างระดับบางโคลัด', 'ที่ตรงข้ามแม็คโคร', 'บางบอน', 'รถชนกัน', '5', 'คัน', 'ขว…', '07:52']
['ถนนกาญจนาภิเษก', 'จาก', 'ต่างระดับเอกชัย', ';ที่บริเวณบนต่างระดับบางโคลัด', 'รถเก๋ง', '2คัน', 'ชนกัน', 'ขวางช่องทางขวา', 'การจราจรเคลื่อนตัวช้า',
'08:40']
['ถนนกาญจนาภิเษก', 'ตรงข้ามหมู่บ้านพระปิ่น3', ';ต่างระดับบางคูเวียง', 'ที่บริเวณตรงข้ามวัดพระเจ็น', 'รถ…', '10:48', 'ถนนกาญจนาภิเษก']
['ต่างระดับลาดหลุมแก้ว', ';ต่างระดับปทุมเสนา', 'ที่กม65+500', 'รถชนกัน', '3คัน', 'ขวางช่องทา…', '09:10', 'ถนนกาญจนาภิเษก', 'ตรงข้ามหมู่บ้านพระปิ่น3']
```

ภาพประกอบ 32 การประมวลผลภาษาธรรมชาติ

จากการวิเคราะห์โครงสร้างของแต่ภาษาไทยและภาษาอังกฤษพบว่า โครงสร้างภาษาอังกฤษนั้นจะมีรูปแบบเป็นที่คำอยู่แล้ว โดยการตัดคำในภาษาอังกฤษจะใช้ย่อหน้าประโยคคำเดี่ยว ๆ หรือช่องว่าระหว่างคำ (space) ในการตัดคำทำให้การตัดคำในข้อความภาษาอังกฤษอย่างเดียวจะได้ผลลัพธ์ของคำที่ออกมาตามพจนานุกรม แต่เมื่อข้อความที่ใช้งานเป็นคำภาษาไทยจะมีการตัดคำที่ซับซ้อนกว่าและคำในภาษาไทยจะมีการเว้นวรรคและมีช่องว่างระหว่างคำ จึงใช้

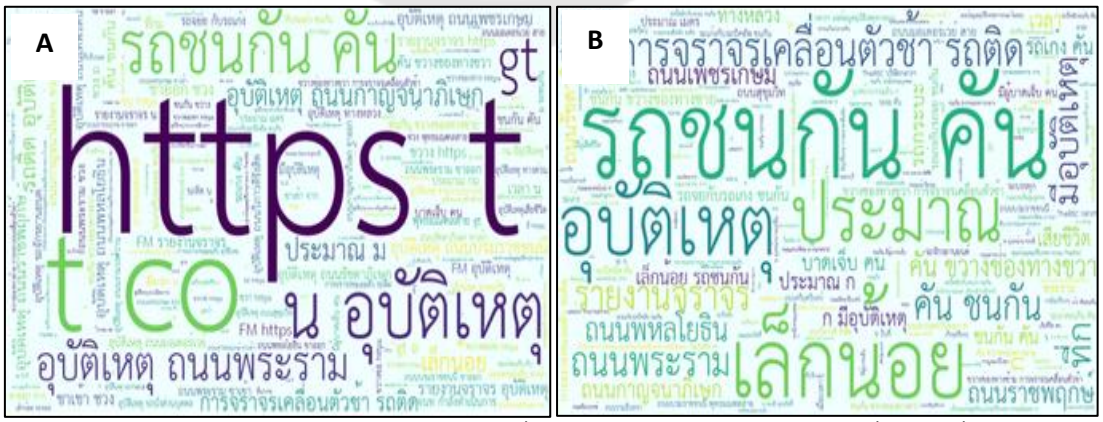
เครื่องมือ whitespace+newline มาช่วยในการแบ่งคำให้มีความถูกต้องมากขึ้น แต่อาจจะมีคำที่ระบบมีความผิดพลาดหรือตัดคำผิด ส่วนมากจะเป็นคำที่มีความหมายได้เมื่อแยกคำ หรือคำที่ระบบทวิตเตอร์ได้ตัดคำนั้นออกไปจากการบันทึก เช่น อ… ไม่มี… จะเป็นคำที่ไม่มีความหมาย ดังภาพประกอบ 34



ภาพประกอบ 33 กระบวนการตัดประโยค

2.2 การทำความสะอาดข้อมูลด้วยนิพจน์ปรกติ (Regular expression)

กระบวนการทำความสะอาดข้อมูลด้วย Regular Expression ช่วยกรองเอาอักษรหรือสัญลักษณ์ที่ไม่ต้องการหรือไม่มีผลในการวิเคราะห์ออกไป ในงานวิจัยนี้ได้ทำการกรองข้อความในชุดข้อมูลให้ออกมามีความสมบูรณ์ที่สุด โดยจะทำให้ได้ผลลัพธ์ในการประมวลผลภาษาธรรมชาติได้ดีมากขึ้น โดยข้อความที่ถูกบันทึกมาจะมีคำต่างๆที่ไม่ต้องการเช่น ลิงค์ของเวปไซต์ ชื่อของบัญชีผู้ใช้งาน จากภาพประกอบ 34 (A) จะมีข้อความที่เราไม่ต้องการนำมาวิเคราะห์ แต่เมื่อทำการกรองข้อมูลด้วย Regular Expression แล้วข้อความที่ไม่ต้องการจะนำไปวิเคราะห์ เช่น ลิงค์ของเว็ปไซต์ ข้อความที่เป็นภาษาไทยหรือภาษาอังกฤษที่ไม่สามารถอธิบายความหมายได้ หรืออักษรพิเศษต่างๆ จะถูกตัดออกไปจากการวิเคราะห์ ดังภาพประกอบ 34 (B)



ภาพประกอบ 34 เปรียบเทียบระหว่างข้อมูลที่ยังไม่ถูกตัดคำ (A) และข้อมูลที่ตัดคำที่ไม่ต้องการด้วย Regular Expression (B)

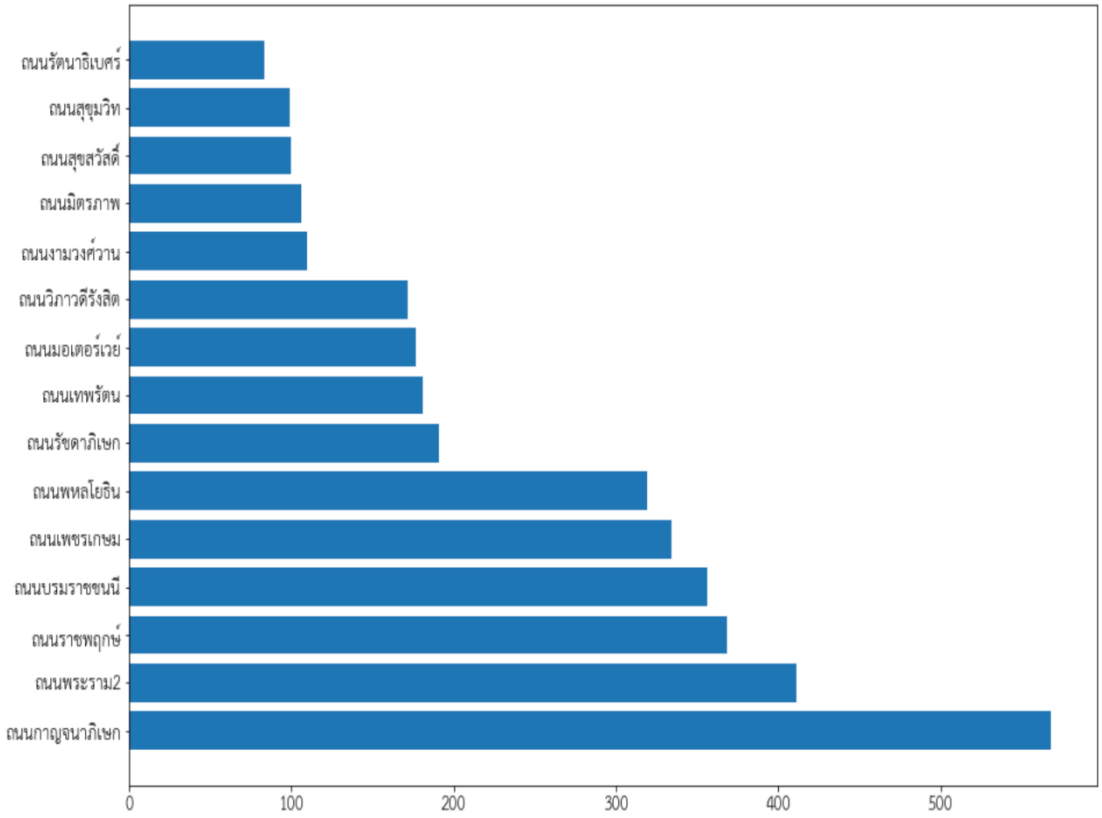
2.4. การจัดทำกลุ่มข้อมูลและความถี่ข้อมูล

การวิเคราะห์ดัชนีถ้อยคำก้อนเมฆ (Word Cloud) เป็นการจัดลำดับข้อมูลเพื่อนำมาแสดงผลให้มีความเข้าใจง่ายขึ้น สามารถมองเห็นความถี่ของข้อความที่ทำการประมวลผลภาษาได้ง่ายขึ้น โดยวิเคราะห์จำนวนความถี่ของคำที่เกิดขึ้นในข้อความนั้นโดยอยู่ในรูปแบบการจัดทำชุดถุงข้อมูลเป็นการสร้างคลังคำศัพท์ (Bag of words) การแปลงคำแต่ละคำที่ไม่ซ้ำกัน มีการนับคำซ้ำเพื่อจัดทำสถิติข้อมูล เช่น ชื่อถนนที่เกิดอุบัติเหตุสูงสุด ดังภาพประกอบ 35

```
Counter({'กาญจนาภิเษก': 467, 'พระราม2': 326, 'บรมราชชนนี': 296, 'ราชพฤกษ์': 283, 'เพชรเกษม': 269, 'พหลโยธิน': 238, 'รัชดาภิเษก': 148,
'มอเตอร์เวย์': 144, 'เทพรัตน': 140, 'วิภาวดีรังสิต': 136, 'งามวงศ์วาน': 88, 'สุขุมวิท': 74, 'มิตรภาพ': 72, 'สุขสวัสดิ์': 69, 'จรัญสนิทวงศ์': 64, 'ร
ัตนาธิเบศร์': 61, 'กัลปพฤกษ์': 60, 'เอกชัยบางบอน': 56, 'ศรีนครินทร์': 56, 'พระราม4': 54, 'พระราม9': 52, 'รามอินทรา': 48, 'ประดิษฐ์มนูธรรม': 4
```

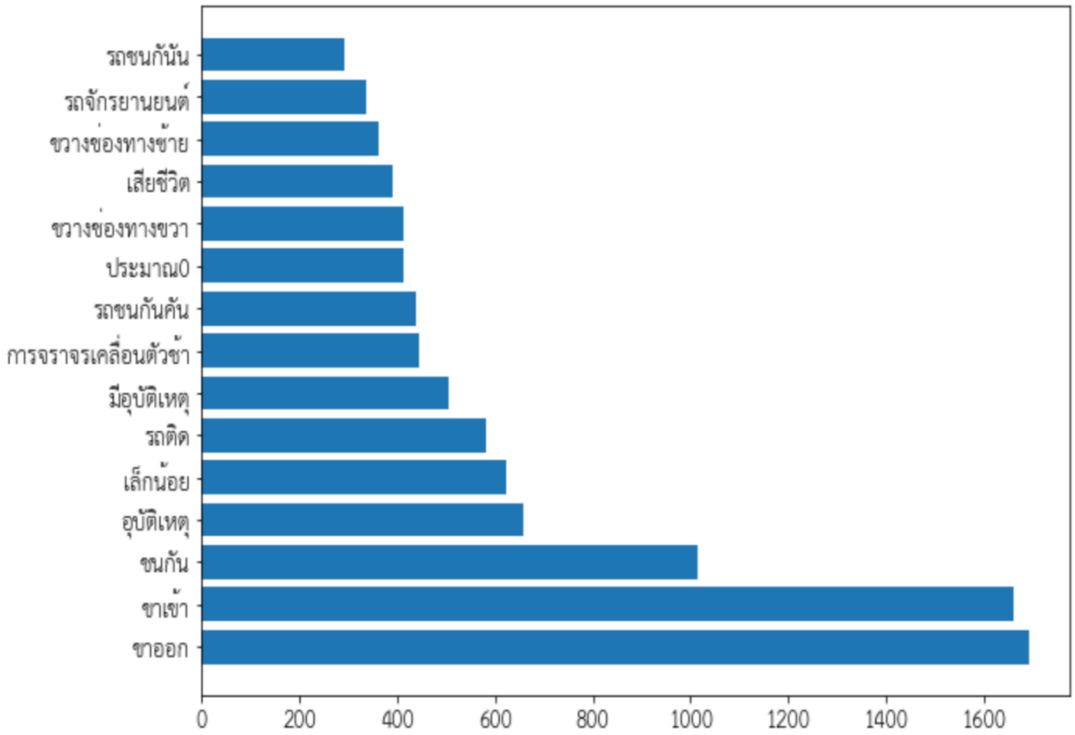
ภาพประกอบ 35 ความถี่ถนนที่มีการเกิดอุบัติเหตุสูงสุดแสดงผลในรูปแบบถ้อยคำก้อนเมฆ

จากกระบวนการประมวลภาษาธรรมชาติข้อมูลที่ถูกกรองแล้วจำนวน 13,095 ข้อมูลที่ได้จัดทำในรูปแบบถ้อยคำก้อนเมฆ (Word Cloud) คำที่มีความถี่ของข้อความสูงที่สุด โดยจำแนกเป็นความถี่ของถนนที่เกิดอุบัติเหตุสูงสุด 15 อันดับแรกที่มีการเกิดอุบัติเหตุสูงสุด ได้แก่ ถนนกาญจนาภิเษก ถนนพระราม2 ถนนบรมราชชนนี ถนนราชพฤกษ์ ถนนเพชรเกษม ถนนพหลโยธิน ถนนรัชดาภิเษก ถนนมอเตอร์เวย์ ถนนวิภาวดี-รังสิต ถนนเทพรัตน์ ถนนงามวงศ์วาน ถนนมิตรภาพ ถนนสุขุมวิท ถนนสุขสวัสดิ์ ถนนจรัญสนิทวงศ์ ตามลำดับ ดังภาพประกอบ 36-37 และจำแนกเป็นความถี่ของข้อความที่เกิดขึ้นสูงสุดได้แก่ ขาเข้า ขาออก อุบัติเหตุ รถชนกัน คันเล็กน้อย รถติด การจราจรเคลื่อนตัวช้า ดังภาพประกอบ 38-39



ภาพประกอบ 36 กราฟแสดงความถี่ของถนนที่เกิดอุบัติเหตุสูงสุด

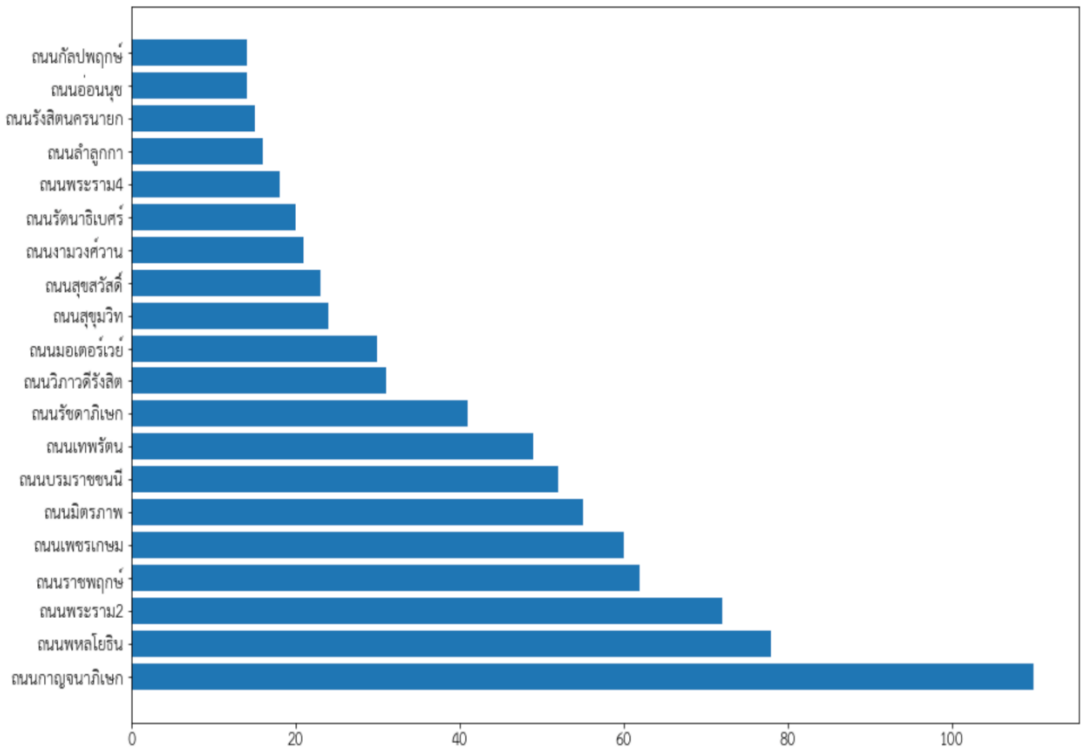
ภาพประกอบ 37 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ถนนที่เกิดอุบัติเหตุสูงสุด



ภาพประกอบ 38 กราฟแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุ

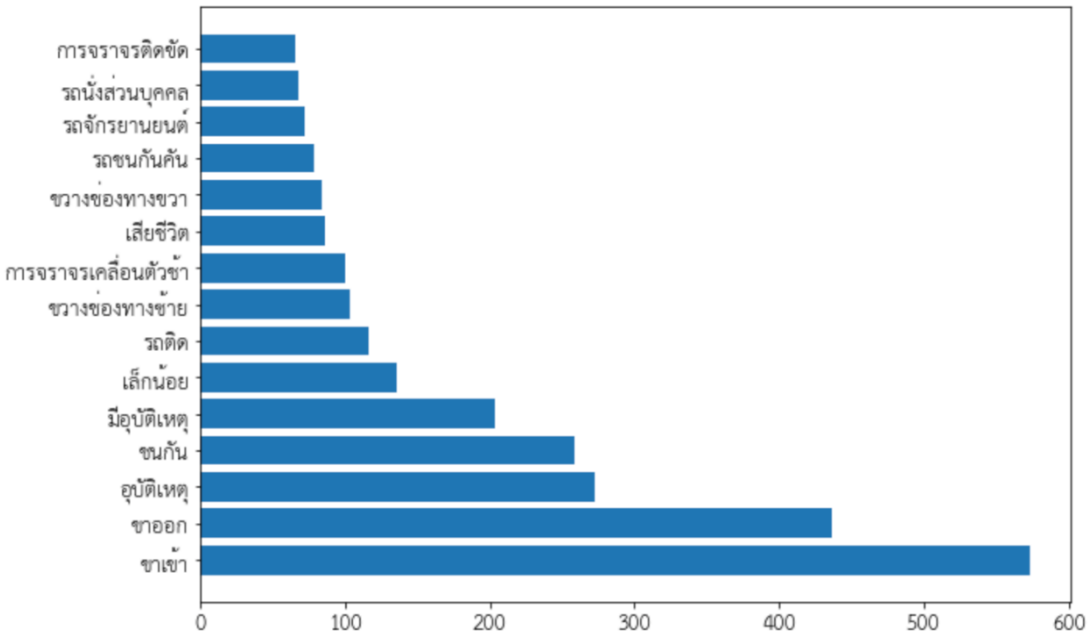
ภาพประกอบ 39 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุ

เมื่อพิจารณาการประมวลผลภาษาธรรมชาติโดยแยกช่วงเวลา เช้า 06:00-09:00 น. มีข้อมูลจำนวน 2,374 ข้อมูล พบว่าความถี่ของถนนที่เกิดอุบัติเหตุสูงสุด 10 อันดับ ได้แก่ ถนนกาญจนาภิเษก มีจำนวนอุบัติเหตุ 110 เหตุการณ์ ถนนพหลโยธิน มีจำนวนอุบัติเหตุ 78 เหตุการณ์ ถนนพระราม2 มีจำนวนอุบัติเหตุ 72 เหตุการณ์ ถนนราชพฤกษ์ มีจำนวนอุบัติเหตุ 62 เหตุการณ์ ถนนเพชรเกษม มีจำนวนอุบัติเหตุ 60 เหตุการณ์ ถนนมิตรภาพ มีจำนวนอุบัติเหตุ 55 เหตุการณ์ ถนนบรมราชชนนี มีจำนวนอุบัติเหตุ 52 เหตุการณ์ ถนนเทพรัตน มีจำนวนอุบัติเหตุ 49 เหตุการณ์ ถนนรัชดาภิเษก มีจำนวนอุบัติเหตุ 31 เหตุการณ์ และถนนวิภาวดีรังสิตมีจำนวนอุบัติเหตุ 31 เหตุการณ์ ดังภาพประกอบ 40-41 และจากการประมวลผลภาษาธรรมชาติมีความถี่ของคำที่เกี่ยวข้องกับอุบัติเหตุ ได้แก่ ขาเข้า ขาออก อุบัติเหตุ ชนกัน รถติด ขวางช่องทางซ้าย ดังภาพประกอบ 41-42



ภาพประกอบ 40 กราฟแสดงความถี่ถนนที่เกิดอุบัติเหตุสูงสุดในช่วงเวลา 06:00 – 09:00 น.

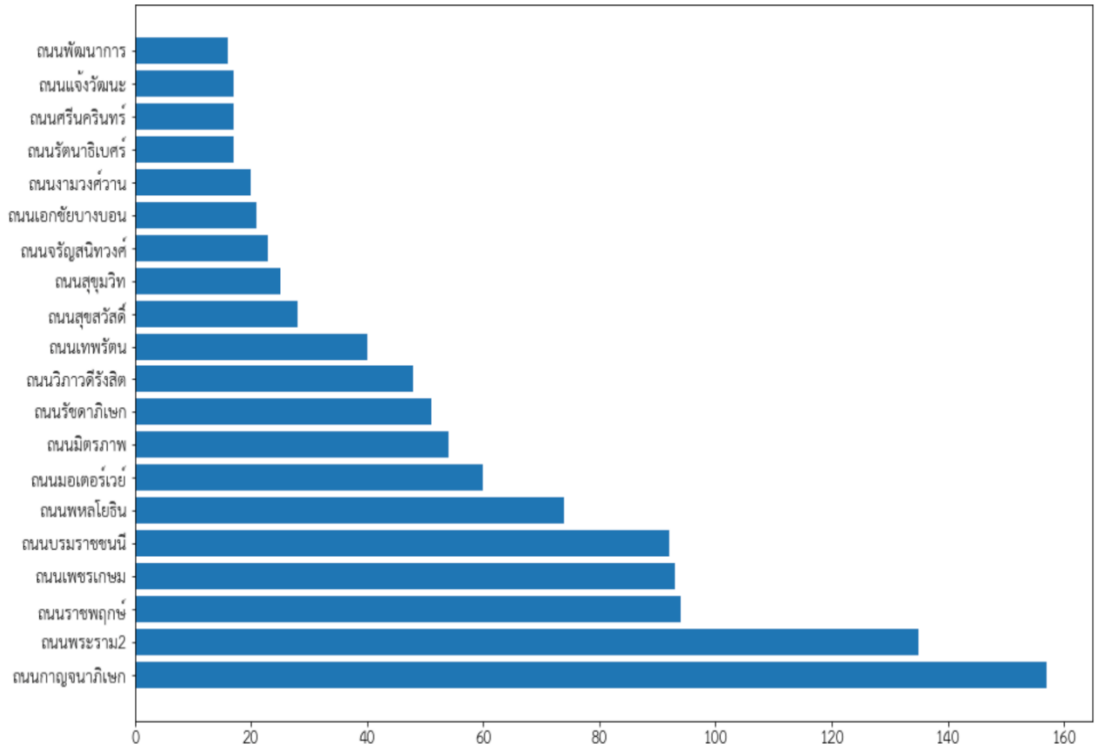
ภาพประกอบ 41 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ถนนที่เกิดอุบัติเหตุสูงสุดในช่วงเวลา 06:00 – 09:00 น.



ภาพประกอบ 42 กราฟแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุในช่วงเวลา 06:00-09:00 น.

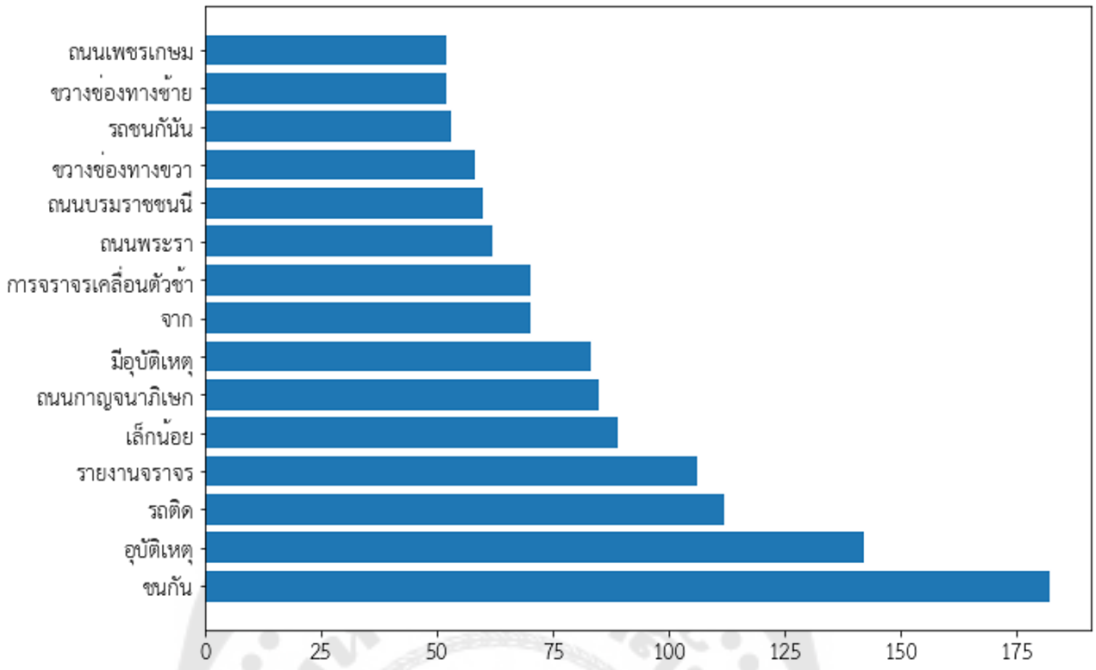
ภาพประกอบ 43 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุในช่วงเวลา 06:00-09:00 น.

เมื่อพิจารณาการประมวลผลภาษาธรรมชาติโดยแยกช่วงเวลากลางวัน 11:00 – 14:00 น. มีข้อมูลจำนวน 2,429 ข้อมูล พบว่ามีความถี่ของถนนที่เกิดอุบัติเหตุสูงสุด 10 อันดับ ได้แก่ ถนนกาญจนาภิเษก มีจำนวนอุบัติเหตุ 157 เหตุการณ์ ถนนพระราม2 มีจำนวนอุบัติเหตุ 135 เหตุการณ์ ถนนราชพฤกษ์ มีจำนวนอุบัติเหตุ 94 เหตุการณ์ ถนนเพชรเกษม มีจำนวนอุบัติเหตุ 93 เหตุการณ์ ถนนบรมราชชนนี มีจำนวนอุบัติเหตุ 92 เหตุการณ์ ถนนพหลโยธิน มีจำนวนอุบัติเหตุ 74 เหตุการณ์ ถนนมอเตอร์เวย์ มีจำนวนอุบัติเหตุ 60 เหตุการณ์ ถนนมิตรภาพ มีจำนวนอุบัติเหตุ 54 เหตุการณ์ ถนนรัชดาภิเษก มีจำนวนอุบัติเหตุ 51 เหตุการณ์ และถนนวิภาวดีรังสิตมีจำนวนอุบัติเหตุ 48 เหตุการณ์ ดังภาพประกอบ 44-45 และจากการประมวลผลภาษาธรรมชาติมีความถี่ของคำที่เกี่ยวข้องกับอุบัติเหตุ ได้แก่ ชนกัน อุบัติเหตุ รถติด ขวางช่องทางขวา ขวางช่องทางซ้าย ดังภาพประกอบ 46-47



ภาพประกอบ 44 กราฟแสดงความถี่ถนนที่เกิดอุบัติเหตุสูงสุดในช่วงเวลา 11:00 – 14:00 น.

ภาพประกอบ 45 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ถนนที่เกิดอุบัติเหตุสูงสุดในช่วงเวลา 11:00 – 14:00 น.

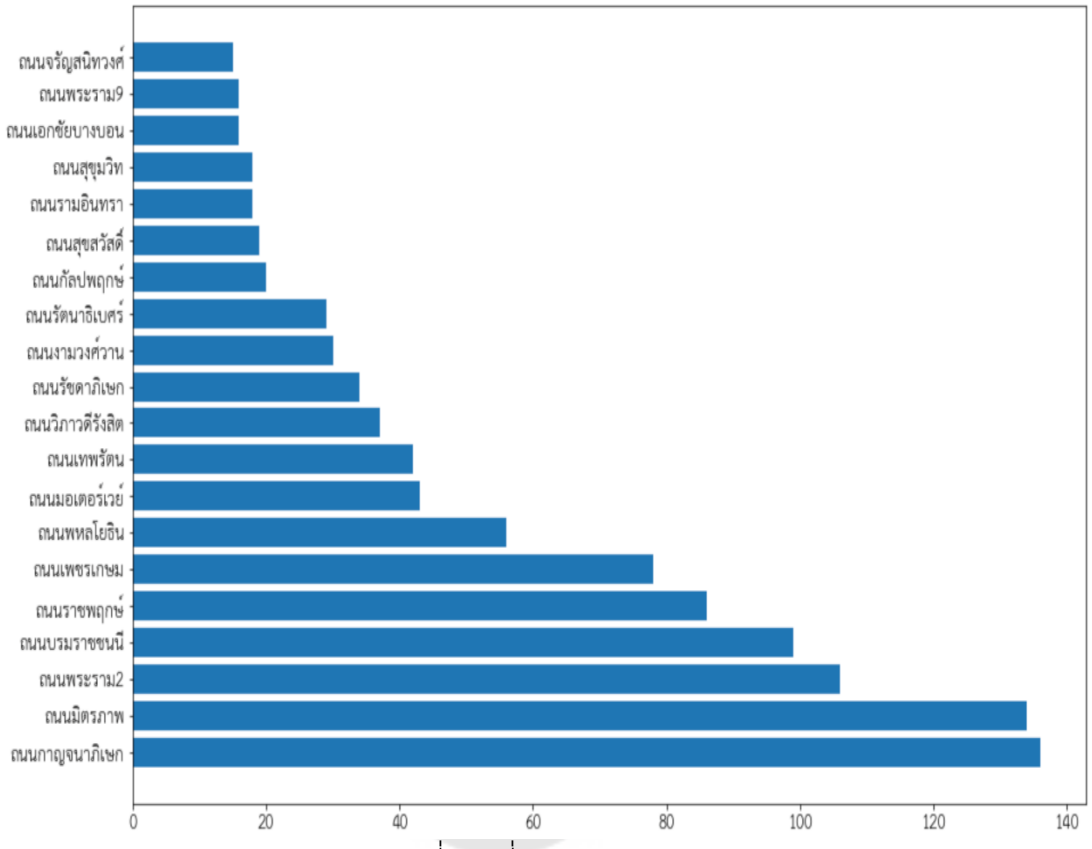


ภาพประกอบ 46 กราฟแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุในช่วงเวลา 11:00-14:00 น.

ภาพประกอบ 47 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุในช่วงเวลา 11:00-14:00 น.

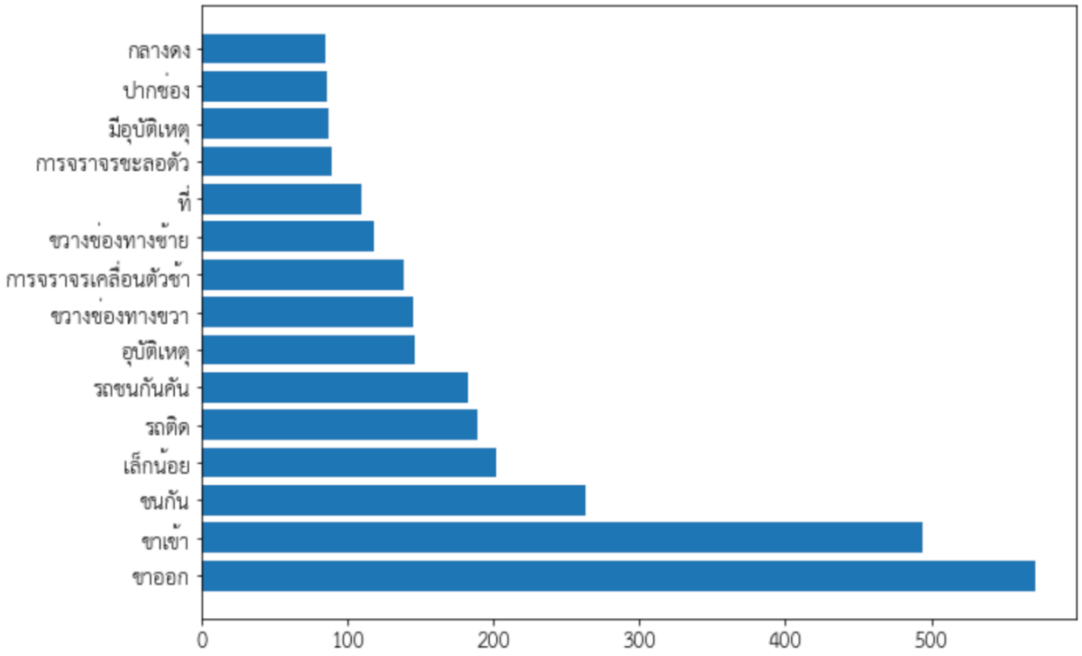
เมื่อพิจารณาการประมวลผลภาษาธรรมชาติโดยแยกช่วงเวลาเย็น 16:00-20:00 น. มีข้อมูลจำนวน 2,752 ข้อมูล พบว่ามีความถี่ของถนนที่เกิดอุบัติเหตุสูงสุด 10 อันดับ ได้แก่ ถนนกาญจนาภิเษก มีจำนวนอุบัติเหตุ 136 เหตุการณ์ ถนนมิตรภาพ มีจำนวนอุบัติเหตุ 134 เหตุการณ์ ถนนพระราม2 มีจำนวนอุบัติเหตุ 106 เหตุการณ์ ถนนบรมราชชนนี มีจำนวนอุบัติเหตุ 99

เหตุการณ์ ถนนราชพฤกษ์ มีจำนวนอุบัติเหตุ 86 เหตุการณ์ ถนนเพชรเกษม มีจำนวนอุบัติเหตุ 78 เหตุการณ์ ถนนพหลโยธิน มีจำนวนอุบัติเหตุ 56 เหตุการณ์ ถนนมอเตอร์เวย์ มีจำนวนอุบัติเหตุ 43 เหตุการณ์ ถนนเทพรัตน มีจำนวนอุบัติเหตุ 42 เหตุการณ์ และถนนวิภาวดีรังสิตมีจำนวนอุบัติเหตุ 37 เหตุการณ์ ดังภาพประกอบ 48-49 และจากการประมวลผลภาษาธรรมชาติมีความถี่ของคำที่เกี่ยวข้องกับอุบัติเหตุ ได้แก่ ชนกัน อุบัติเหตุ รถติด ขวางช่องทางขวา ขวางช่องทางซ้าย ดังภาพประกอบ 50-51



ภาพประกอบ 48 กราฟแสดงความถี่ถนนที่เกิดอุบัติเหตุสูงสุดในช่วงเวลา 16:00 – 20:00 น.

ภาพประกอบ 49 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ถนนที่เกิดอุบัติเหตุสูงสุดในช่วงเวลา 16:00 – 20:00 น.



ภาพประกอบ 50 กราฟแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุในช่วงเวลา 16:00-20:00 น.

ภาพประกอบ 51 ดัชนีถ้อยคำก้อนเมฆแสดงความถี่ของข้อความที่เกี่ยวกับอุบัติเหตุในช่วงเวลา 16:00-20:00 น.

3. การวิเคราะห์เหมืองข้อมูลกับสถิติการเกิดอุบัติเหตุจากกระทรวงคมนาคม

ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงคมนาคม มีการบันทึกสถิติข้อมูลของอุบัติเหตุในกรุงเทพมหานครและปริมณฑล ปี 2564 จำนวน 4604 เหตุการณ์ โดยถนนที่มีการรายงานอุบัติเหตุสูงสุดคือ ถนนกาญจนาภิเษก จำนวน 758 เหตุการณ์ ถนนพระราม2 จำนวน 150 เหตุการณ์ ถนนบรมราชชนนี จำนวน 106 เหตุกาณ์ โดยได้ทำการบันทึกข้อมูลในรูปแบบจุด (Point) ที่เป็นค่าพิกัดทางภูมิศาสตร์ ดังภาพประกอบ 52

ภาพประกอบ 52 สถิติการเกิดอุบัติเหตุ ปี 2564 ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงคมนาคม

จากข้อมูลจากหน่วยงานสามารถแสดงให้เห็นถึงข้อมูลที่ทางหน่วยงานภาครัฐได้ทำการบันทึกและสามารถเปรียบเทียมกับข้อมูลที่บันทึกจากสื่อสังคมออนไลน์ ที่มีจำนวนการเกิดอุบัติเหตุบนท้องถนนที่มีความถี่สอดคล้องกันกับการบันทึกจากสื่อสังคมออนไลน์ (สำนักงานปลัดกระทรวงคมนาคม, 2564)

4. การวิเคราะห์เหมืองข้อมูลและการหาความสัมพันธ์ระหว่างข้อมูลการเกิดอุบัติเหตุด้วยอัลกอริทึมอพริโอริ (Association Rule Mining using Apriori Algorithm)

การวิเคราะห์หากฎความสัมพันธ์ใช้เทคนิคทำเหมืองข้อมูลคือกระบวนการกระทำกับข้อมูลเพื่อค้นหารูปแบบ กฎ และความสัมพันธ์ของข้อมูลที่อยู่ในชุดข้อมูลนั้นโดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์ เป็นต้น การหาความสัมพันธ์ของข้อมูลข้อมูลชุดนี้ถูกนำไปวิเคราะห์เพื่อสร้างกฎความสัมพันธ์ด้วยอัลกอรึทึม Apriori การสร้างกฎได้กำหนดค่าสนับสนุนขั้นต่ำ และค่าความเชื่อมั้นขั้นต่ำไว้ โดยการกำหนดค่าสนับสนุนต่ำสุดที่ 0.08 เนื่องจากการหาความสัมพันธ์ระหว่างข้อความภาษาไทยจะมีความซับซ้อนในกระบวนการประมวลผลภาษาธรรมชาติมากกว่า จะเห็นได้ว่าอุบัติเหตุบนท้องถนนส่วนใหญ่ คือเหตุการณ์ “รถชนกัน” โดยความสัมพันธ์ระหว่างถนนและอุบัติเหตุ กล่าวคือการเกิดอุบัติเหตุบนถนนมีความสัมพันธ์กับลักษณะของการเกิดอุบัติเหตุและมีความสัมพันธ์กับสถานที่ใกล้เคียงของอุบัติเหตุ เช่น ถนนกาญจนาภิเษก มีความเชื่อมโยงกับคำว่า รถชนกัน มีค่าความเชื่อมั่น 0.15% มีคำที่อธิบายถึงความเสียหายหรือบริเวณจุดเกิดอุบัติเหตุ เช่นคำว่า ต่างระดับเอกชัย มีค่าความเชื่อมั่น 0.12% จะเห็นได้ว่าการรายงานข้อมูลผ่านสื่อสังคมออนไลน์มีการกล่าวถึงคำที่เกี่ยวข้องกับอุบัติเหตุที่เกิดขึ้น

เมื่อทำการหาความสัมพันธ์ระหว่างข้อความจากการประมวลผลภาษาธรรมชาติ จะได้ผลลัพธ์ออกมาเป็นค่าความเชื่อมั่น โดยกำหนดความสัมพันธ์ระหว่างถนนที่มีความถี่ของการเกิดอุบัติเหตุสูงสุด และคำว่าอุบัติเหตุผลของการหาความสัมพันธ์ระหว่างข้อความจะได้ค่าความเชื่อมั่น ดังตารางที่ 6

ตาราง 6 : คำที่เกี่ยวข้องกับคำสำคัญ (keyword) และค่ากฎความสัมพันธ์ของคำสำคัญ (associate's rule)

ถนนกาญจนาภิเษก	ค่าความเชื่อมั่น
จาก	0.128975
รถชนกัน	0.151943
เล็กน้อย	0.121908
ต่างระดับเอกชัย	0.118375

ถนนพระราม2	ค่าความเชื่อมั่น
รถชนกัน	0.155556
ชนกัน	0.145679
ประมาณ	0.138272
คัน	0.135802

ต่างระดับบางโคลัด	0.098940
ประมาณ	0.090106
ต่างระดับเพชรเกษม	0.083039

แสมดำ	0.128395
แยกบ้านแพ้ว	0.123457
จาก	0.113580
เล็กน้อย	0.111111
ต่างระดับพระราม2	0.108642

ถนนบรมราชชนนี	ค่าความเชื่อมั่น
พุทธมณฑลสาย2	0.228169
รถชนกัน	0.228169
เล็กน้อย	0.208451
ประมาณ	0.183099
คัน	0.174648
พุทธมณฑลสาย4	0.166197
คัน,รถชนกัน	0.112676
ชนกัน	0.109859

ถนนราชพฤกษ์	ค่าความเชื่อมั่น
รถชนกัน	0.189041
แยกสวนผัก	0.147945
คัน	0.145205
ชนกัน	0.115068
เล็กน้อย	0.112329
แยกบางขุนกอง	0.109589
จาก	0.101370
ต่างระดับชัยพฤกษ์	0.101370
ต่างระดับบางรักน้อย	0.101370

ถนนเพชรเกษม	ค่าความเชื่อมั่น
ชนกัน	0.206587
เล็กน้อย	0.161677
รถชนกัน	0.155689

ถนนพหลโยธิน	ค่าความเชื่อมั่น
จาก	0.114286
รถชนกัน	0.133333
เล็กน้อย	0.092063

คัน	0.143713
นครปฐม	0.107784
จาก	0.098802
ประมาณ	0.089820
คัน,รถชันกัน	0.089820
การจราจรเคลื่อนตัวช้า	0.083832

ชนกัน	0.088889
รังสิต	0.082540

ถนนรัชดาภิเษก	ค่าความเชื่อมั่น
แยกรัชดาลาดพร้าว	0.105263
รถชนกัน	0.142105
รถติด	0.142105
จาก	0.131579
ชนกัน	0.126316
แยกรัชโยธิน	0.115789
ขวางช่องทางซ้าย	0.110526
คัน	0.110526
การจราจรเคลื่อนตัวช้า	0.100000

ถนนมอเตอร์เวย์	ค่าความเชื่อมั่น
สาย7	0.276596
ต่างระดับทับช้าง	0.212766
รถชนกัน	0.196809
สาย9	0.196809
9, สาย	0.148936
สาย, 7	0.138298

ถนนวิภาวดี-รังสิต	ค่าความเชื่อมั่น
แยกบางเขน	0.116959
จาก	0.181287
รถชนกัน	0.157895

ถนนเทพรัตน์	ค่าความเชื่อมั่น
(บางนาตราด)	0.232044
ชนกัน	0.171271
รถชนกัน	0.149171

ชนกัน	0.146199
คัน	0.128655
แยกหลักสี่	0.128655
เล็กน้อย	0.116959
สนามบินดอนเมือง	0.105263

คัน	0.116022
แยกบางนา	0.110497
ตรงข้ามสนามบินสุวรรณภูมิ	0.099448
สะพานกลับรถ	0.093923
(กิ่งแก้วบางพลี)	0.088398
บางโฉลง	0.082873
คัน,รถชนกัน	0.082873

ถนนงามวงศ์วาน	ค่าความเชื่อมั่น
แยกพงษ์เพชร	0.366972
แยกบางเขน	0.339475
คัน	0.275229
ชนกัน	0.238532
รถติด	0.229358
รถชนกัน	0.211009
แยกแคราย	0.201835
การจราจรเคลื่อนตัวช้า	0.192661
ขวางช่องทางขวา	0.183486

ถนนมิตรภาพ	ค่าความเชื่อมั่น
นครราชสีมา	0.524272
ปากช่อง	0.310680
นครราชสีมา,ปากช่อง	0.271845
สระบุรี	0.252427
กรุงเทพฯ	0.165049
กลางดง	0.155340
มวกเหล็ก	0.135922
ชนกัน	0.116505
นครราชสีมา,กรุงเทพฯ	0.106796

ถนนสุขุมวิท	ค่าความเชื่อมั่น
ชนกัน	0.279661
ขวางช่องทางซ้าย	0.169492

ถนนสุขสวัสดิ์	ค่าความเชื่อมั่น
จาก	0.151515
แยกบางปะแก้ว	0.151515

การจราจรเคลื่อนตัวช้า	0.161017
รถติด	0.152542
การจราจรเคลื่อนตัวช้า,รถติด	0.127119
ขวางช่องทางกลาง	0.118644
บาดเจ็บ	0.110169
รถชนกัน	0.093220
ขวางช่องทางขวา	0.084746

ชนกัน	0.121212
วงเวียนหอนาฬิกา	0.121212
คัน	0.111111
ทางขึ้นด่วนบางพลีสุขสวัสดิ์	0.111111
ทางขึ้นทางด่วนบางพลีสุขสวัสดิ์	0.101010
ขวางช่องทางซ้าย	0.090909
แยกวัดสน	0.090909

ถนนจรัญสนิทวงศ์	ค่าความเชื่อมั่น
แยกบางพลัด	0.337662
แยกบรมราชนนี	0.259740
สะพานพระราม7	0.207792
รถชนกัน	0.181818
สะพานพระราม7,แยกบางพลัด	0.168831
เล็กน้อย	0.155844
ขวางช่องทางซ้าย	0.129870
คัน	0.129870
ชนกัน	0.129870

ถนนกัลปพฤกษ์	ค่าความเชื่อมั่น
กาญจนาภิเษก	0.112676
แยกกำนันแม้น	0.535211
แยกราชพฤกษ์	0.450704
แยกราชพฤกษ์,แยกกำนันแม้น	0.338028
รถชนกัน	0.169014
คัน	0.154930
จาก	0.140845
ประมาณ	0.126761
แยกสุขาภิบาล1	0.126761

ถนนรัตนาธิเบศร์	ค่าความเชื่อมั่น
จาก	0.182927
แยกแคราย	0.182927

ถนนเอกชัยบางบอน	ค่าความเชื่อมั่น
ชนกัน	0.287879
ต่างระดับเอกชัย	0.196970

รถชนกัน	0.170732
คัน	0.158537
ชนกัน	0.158537
เล็กน้อย	0.146341
เลี่ยงเมืองนนท์	0.134146
แยกบางพลู	0.134146
ถนนเลี่ยงเมืองนนท์	0.121951

บางบอน3	0.166670
บางบอน1	0.151515
วัดโพธิ์แจ้	0.151515
บางบอน5	0.136364
ประมาณ	0.136364

ถนนศรีนครินทร์	ค่าความเชื่อมั่น
แยกศรีนุช	0.176471
แยกพัฒนาการ	0.132353
ชนกัน	0.102941
สะพานลอยวัดศรีเอี่ยม	0.102941
จาก	0.102941
การจราจรเคลื่อนตัวช้า	0.088235
คัน	0.088235
เล็กน้อย	0.088235
แยกลาซาล	0.088235

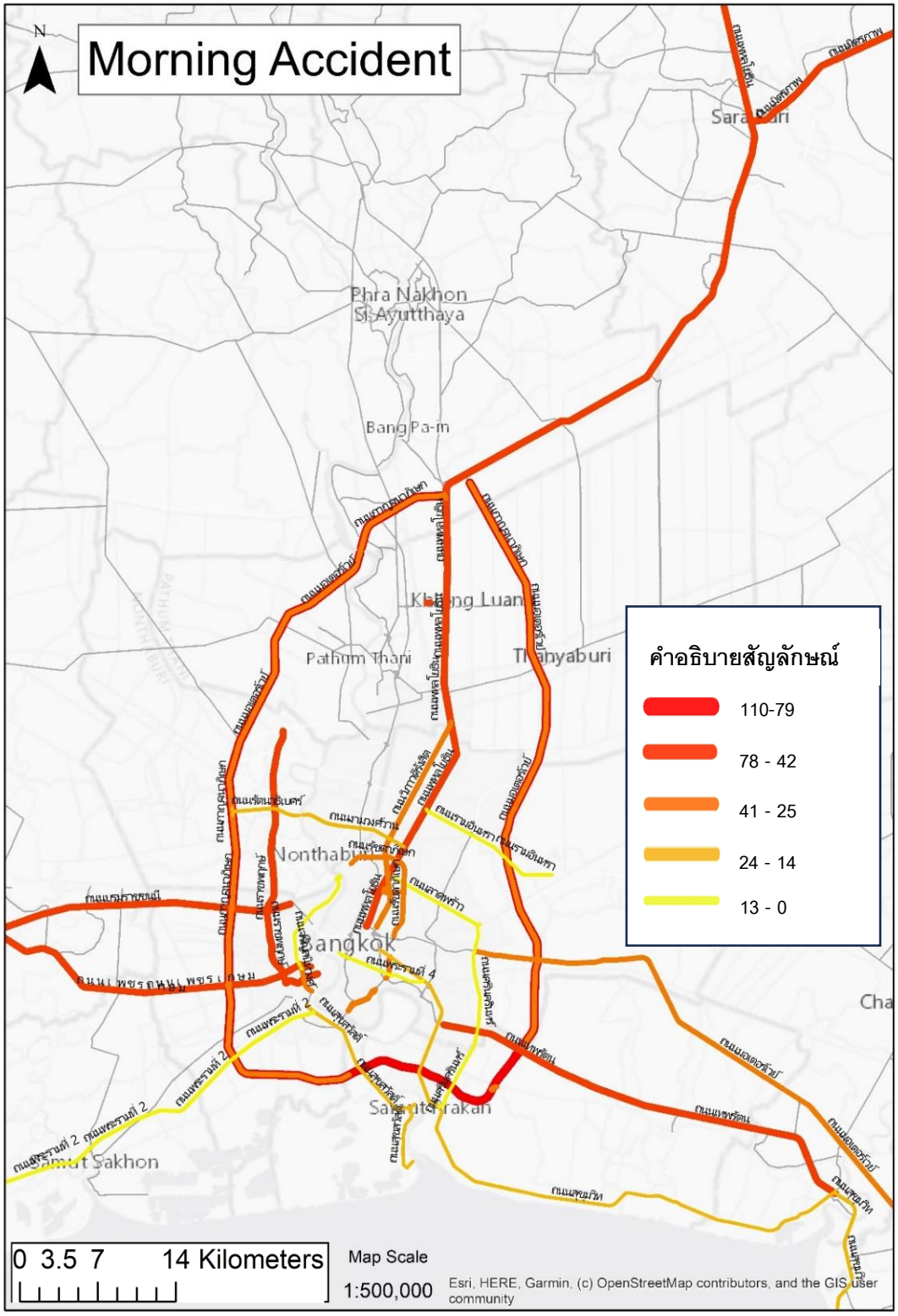
ถนนพระราม4	ค่าความเชื่อมั่น
แยกกล้วยน้ำไท	0.313433
แยกเกษมราษฎร์	0.313433
ชนกัน	0.223881
แยกกล้วยน้ำไท,แยกเกษมราษฎร์	0.208955
ขวางช่องทางขวา	0.164179
ใต้ทางด่วนพระราม4	0.164179
เล็กน้อย	0.134328
แยกถนนวิทยุ	0.104478
จาก	0.089552

จากการหาความสัมพันธ์ของข้อความ โดยใช้เทคนิคอัลกอริทึมอพริโอริ ทำให้ทราบถึงความเชื่อมโยงกับการเกิดอุบัติเหตุบนถนน ข้อมูลถนนที่นำมาสร้างกฎคือ รายชื่อถนนที่มีการรายงานอุบัติเหตุสูงสุด มีค่าความเชื่อมั่นตั้งแต่ 0.08 – 0.52 จำนวน 20 ถนนที่มีความถี่สูงสุด

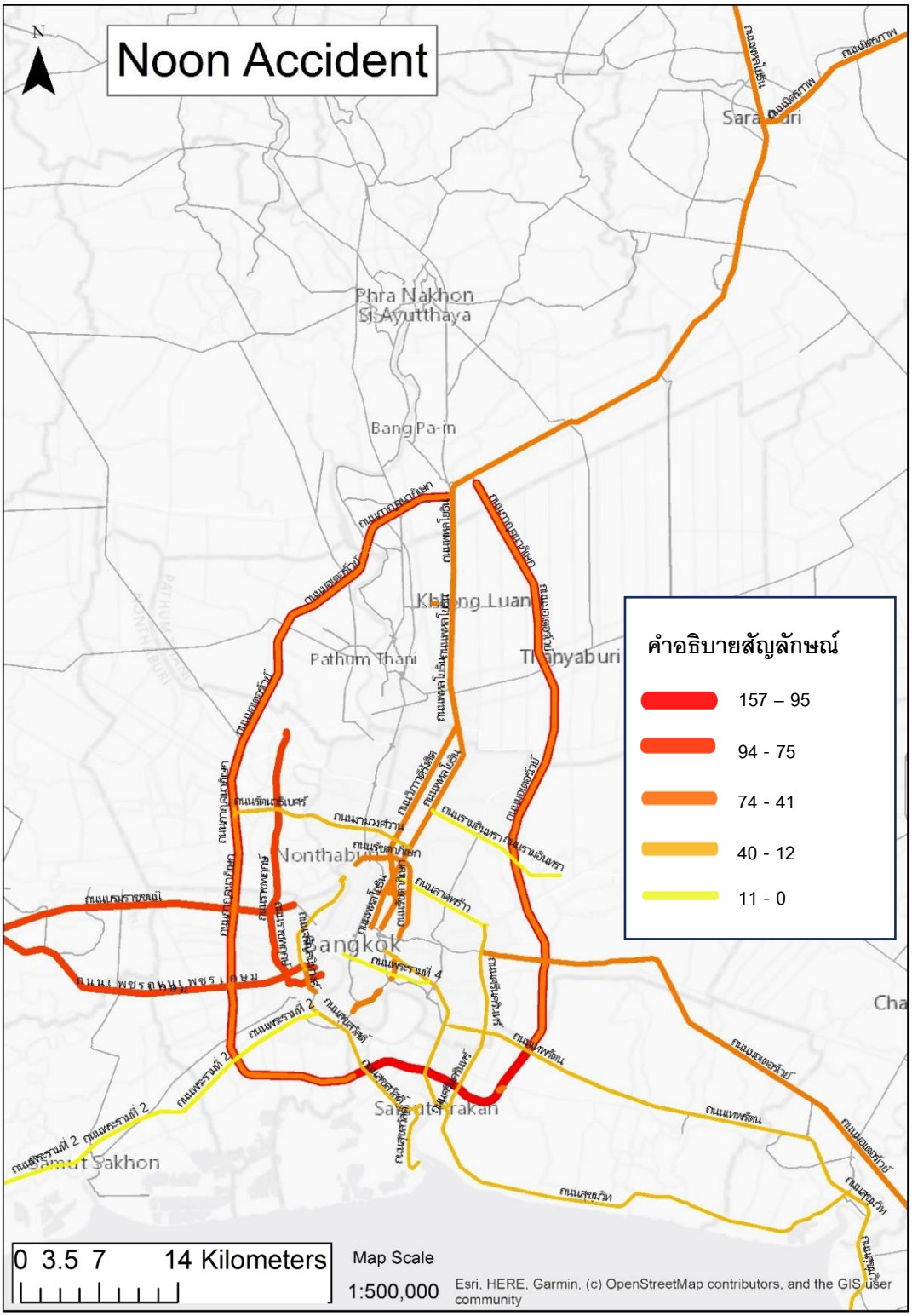
ได้แก่ ถนนกาญจนาภิเษก ถนนพระราม2 ถนนบรมราชนนี ถนนราชพฤกษ์ ถนนเพชรเกษม ถนนพหลโยธิน ถนนรัชดาภิเษก ถนนมอเตอร์เวย์ถนนวิภาวดี-รังสิต ถนนเทพรัตน์ ถนนงามวงศ์วาน ถนนมิตรภาพ ถนนสุขุมวิท ถนนสุขสวัสดิ์ ถนนจรัญสนิทวงศ์ ถนนกัลปพฤกษ์ ถนนรัตนาธิเบศร์ ถนนเอกชัยบางบอน ถนนศรีนครินทร์ ถนนพระราม4 ตามลำดับ โดยคำที่มีความสัมพันธ์กับชื่อถนน และ อุบัติเหตุ คำที่มีความสัมพันธ์จะเป็นคำที่มีการอธิบายถึงสภาพการเกิดอุบัติเหตุ เช่น คือ คำว่า “รถชนกัน” “ชนกัน คำที่อธิบายเพิ่มเติมถึงพื้นที่การเกิดเหตุ เช่น คำว่า “ต่างระดับเอกชัย” “ต่างระดับทับช้าง” “แยกรัชโยธิน” คำที่เป็นการบ่งบอกถึงรายละเอียดเพิ่มเติมของเหตุการณ์ เช่น “ประมาณ” “จาก” “เล็กน้อย” “รถติด” “ขวางช่องทางขวา” คำที่อธิบายถึงจำนวนของยานพาหนะที่เกิดอุบัติเหตุ เช่น “คัน” เป็นต้น

5. ตำแหน่งของข้อมูลการเกิดอุบัติเหตุด้วยการแปลงข้อความเป็นระบบพิกัดทางภูมิศาสตร์ (Location of Accident using Geocoding)

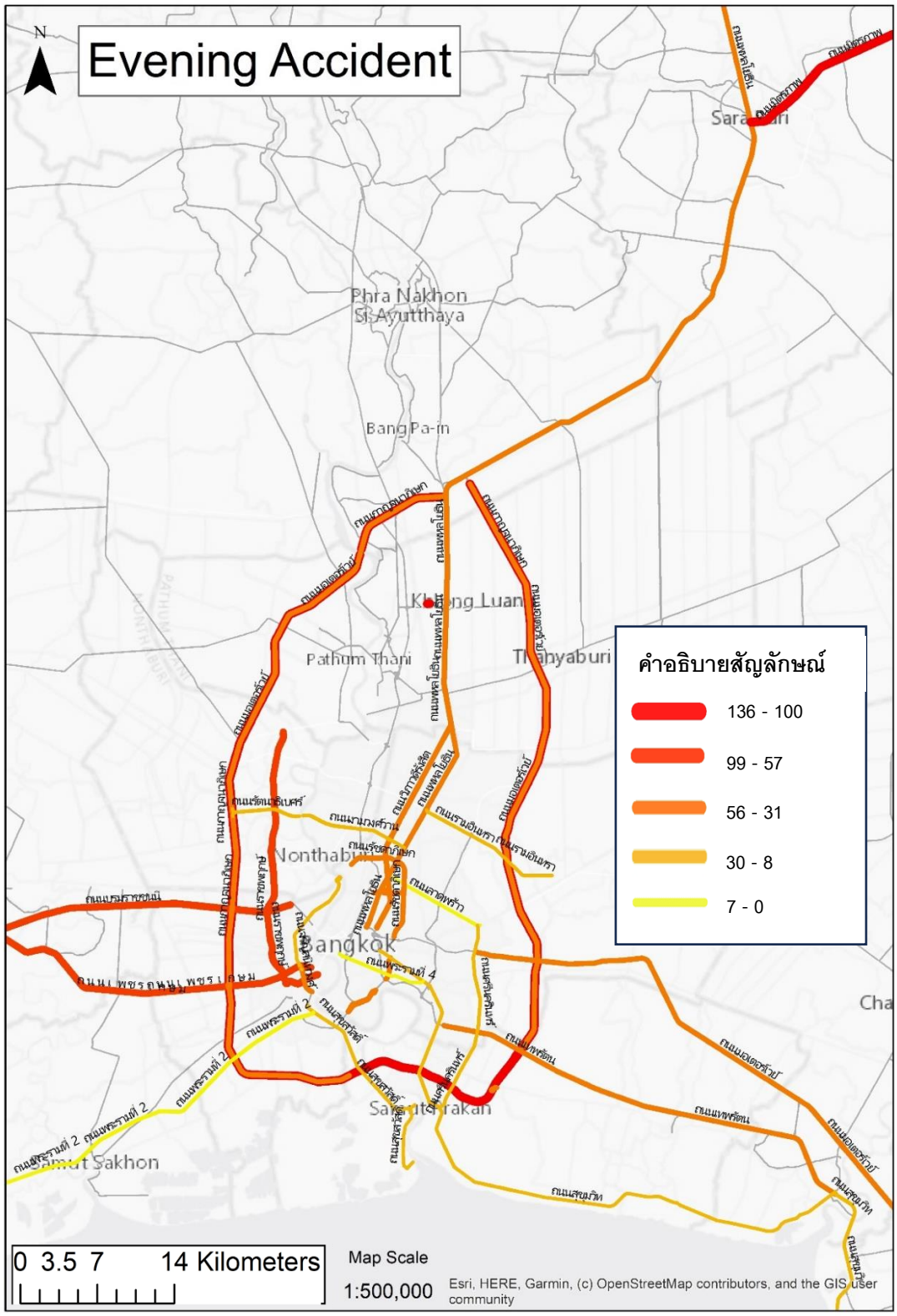
ผลการศึกษากระบวนการเข้ารหัสจากข้อความให้เป็นข้อมูลเชิงพื้นที่ (Geocoding) พบว่ามีอุบัติเหตุที่รายงานทั้งหมด 7555 เหตุการณ์ที่อยู่ในถนนที่มีความถี่ในการเกิดอุบัติเหตุสูงสุด 15 อันดับแรกที่ได้บันทึกจากบัญชี Js100radio และ FM91Radiopro โดยนำไปประมวลผลภาษาธรรมชาติ และหาความสัมพันธ์ของข้อความ โดยแบ่งสถิติของข้อมูลอุบัติเหตุตามช่วงเวลาเร่งด่วน ได้แก่ ช่วงเวลาเช้า 06:00-09:00 น. เวลากลางวัน 11:00-14:00 น. เวลาเย็น 16:00-20:00 น. มีจำนวนอุบัติเหตุ 2,374 2,429 และ 2,752 เหตุการณ์ตามลำดับถนนที่มีการเกิดอุบัติเหตุสูงสุดในทุกช่วงเวลาคือ ถนนกาญจนาภิเษก ถนนพระราม2 และถนนบรมราชนนี มีจำนวนอุบัติเหตุทั้งหมด 403 347 และ 272 เหตุการณ์ ตามลำดับ อุบัติเหตุที่เกิดขึ้นในแต่ละถนนเนื่องจาก การใช้งานที่แตกต่างกันในแต่ละช่วงเวลา จากการสกัดข้อความจาก Twitter เป็นรายชื่อถนน และสร้างแผนที่เส้นทางความถี่ของความหนาแน่นการจราจรและการเกิดอุบัติเหตุ โดยทำแผนที่แบ่งตามลักษณะการกระจายของข้อมูล โดยการแบ่งชั้นข้อมูลความหนาแน่นด้วยวิธีการแบ่งกลุ่มของข้อมูลแบบธรรมชาติ (Natural break) พบว่าถนนที่มีเส้นสีแดงเป็นถนนที่มีการรายงานอุบัติเหตุสูงสุดและถนนที่มีอุบัติเหตุรองลงมาจะแสดงด้วยระดับสีที่อ่อนลง แสดงให้เห็นถึงลักษณะความแตกต่างของความถี่ของถนนที่ต่างกันในแต่ละช่วงเวลา ดังภาพประกอบ 53 – 55 ซึ่งมีความสอดคล้องกับรายงานถนนที่มีการจราจรหนาแน่น ติดขัด ของหน่วยงานภาครัฐที่ได้ทำการเก็บข้อมูลเอาไว้ โดยสาเหตุของการจราจรที่หนาแน่น เป็นผลมาจากอุบัติเหตุที่เกิดขึ้นด้วย เพราะเมื่อเกิดอุบัติเหตุจะทำให้การสัญจรเป็นไปได้ยากขึ้น จากการที่มีรถที่เกิดอุบัติเหตุได้ขวางหรือปิดช่องทางการจราจร



ภาพประกอบ 53 ถนนที่มีความถี่ของการเกิดอุบัติเหตุสูงสุดในช่วงเช้า 06.00 ถึง 09.00



ภาพประกอบ 54 ถนนที่มีความถี่ของการเกิดอุบัติเหตุสูงสุดในช่วงกลางวัน 11.00 ถึง 14.00



ภาพประกอบ 55 ถนนที่มีความถี่ของการเกิดอุบัติเหตุสูงสุดในช่วงเย็น 16:00-20:00

บทที่ 5
สรุปผล อภิปรายผล และข้อเสนอแนะ

การศึกษากระบวนการประมวลผลภาษาธรรมชาติ จากข้อมูลสื่อสังคมออนไลน์เพื่อการคาดการณ์อุบัติเหตุเชิงพื้นที่ ในช่วงเวลา พ.ศ.2564 – 2565 โดยมีวัตถุประสงค์เพื่อพัฒนาระบบฐานข้อมูลเชิงพื้นที่จากข้อมูลสื่อสังคมออนไลน์ วิเคราะห์ข้อมูลที่เกี่ยวข้องกับอุบัติเหตุและประมาณค่าเชิงเส้นทางการเกิดอุบัติเหตุจากการประมวลผลภาษาธรรมชาติ ประกอบด้วยหัวข้อดังนี้

1 สรุปผล
2 อภิปรายผล
3 ข้อเสนอแนะ

1 สรุปผล

การประมวลผลภาษาธรรมชาติเชิงตำแหน่งผ่านสื่อสังคมออนไลน์ กรณีศึกษาแพลตฟอร์ม Twitter โดยแบ่งการวิเคราะห์เป็น 3 ส่วน ได้แก่ การพัฒนาคลังฐานข้อมูลเชิงพื้นที่ในการประมวลผลภาษาธรรมชาติ รวบรวมข้อมูลที่ใช้ในการพัฒนาระบบฐานข้อมูลและออกแบบระบบฐานข้อมูลและหาความสัมพันธ์ของข้อความ ตลอดจนการประมวลผลภาษาธรรมชาติและการวิเคราะห์เพื่อหาความสัมพันธ์ระหว่างข้อมูลเชิงพื้นที่และเวลา มีการใช้เครื่องมือเพื่อช่วยในการตัดคำในรูปแบบภาษาไทย เนื่องจากภาษาไทย มีการแบ่งคำและวรรณยุกต์ที่ยากและซับซ้อนกว่าภาษาอังกฤษ จึงเลือกใช้โมดูล PythaiNLP และเครื่องมือ Newmm engine เพื่อช่วยในการตัดคำให้มีความถูกต้องและแม่นยำ และการใช้เครื่องมือ Regular Expression เพื่อกรองคำที่ไม่สมบูรณ์และไม่ต้องการนำมาวิเคราะห์ เพื่อให้ข้อมูลมีความถูกต้องมากที่สุด และนำคำที่ตัดได้ไปหาความถี่ของข้อความและหาความสัมพันธ์ระหว่างข้อความด้วยอัลกอริทึมอพริโอริ และการระบุตำแหน่งพิกัดทางภูมิศาสตร์ถูกนำมาเป็นข้อมูลสำคัญในการวิเคราะห์การเกิดอุบัติเหตุบนท้องถนนจากการรายงานผ่านระบบสื่อสังคมออนไลน์ ผลการศึกษาการวิเคราะห์เหมืองดัชนีข้อมูลเชิงตำแหน่งพบว่า ลักษณะอุบัติเหตุในแต่ละช่วงเวลามีลักษณะการเกิดในบริเวณที่แตกต่างกัน โดยช่วงเช้าเกิดในบริเวณเมืองที่มีความหนาแน่นของการจารจรในลักษณะการเดินทางช่วงเวลาทำงาน ในช่วงกลางวันเกิดบนถนนสายหลักที่เป็นศูนย์กลางของการแยกไปในถนนสายรอง และ

ช่วงเย็นเกิดในถนนที่เป็นการเดินทางระยะไกลมีการขับขี่ข้ามจังหวัดที่อาจจะมีแสงสว่างไม่เพียงพออาจเป็นสาเหตุของการเกิดอุบัติเหตุ จากการเปรียบเทียบข้อมูลจากศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงคมนาคม ที่มีการบันทึกข้อมูลอุบัติเหตุในแต่ละช่วงเวลาและกรมทางหลวงที่มีการบันทึกข้อมูลรายชื่อถนนที่เกิดอุบัติเหตุ พบว่าข้อมูลอุบัติเหตุจราจรจากข้อมูลสื่อสังคมออนไลน์ มีความสอดคล้องกับสถิติข้อมูลจากหน่วยงานที่เก็บบันทึกในปี 2564 สามารถใช้ข้อมูลจากสื่อสังคมออนไลน์เป็นเครื่องมือในการวิเคราะห์การเกิดอุบัติเหตุบนท้องถนน เพื่อใช้ในการวางแผนการปรับปรุงระบบคมนาคมให้มีประสิทธิภาพต่อไปได้ ผลการสร้างกฎความสัมพันธ์พบว่ากฎความสัมพันธ์ที่ได้จากอัลกอริทึม Apriori มีค่าความเชื่อมั่นมากกว่าร้อยละ 90% จากการกำหนดค่าความเชื่อมั่น ทำให้ทราบผลว่าการเกิดอุบัติเหตุบนท้องถนนที่มีความถี่สูงสุด และเหตุการณ์ที่เกิดขึ้นสูงสุด หรือประเภทรถที่เกิดอุบัติเหตุ จำนวนรถที่เกิดเหตุ ความเสียหายที่เกิดขึ้น

2 อภิปรายผล

2.1 การบันทึกข้อมูลจากสื่อสังคมออนไลน์เพื่อนำมาทำเป็นฐานข้อมูล

การบันทึกข้อมูลจากสื่อสังคมออนไลน์ แพลทฟอร์มทวิตเตอร์ จากบัญชี Js100radio และ FM91traffic มีการบันทึกข้อความที่ถูกทวีต และจัดทำระบบฐานข้อมูล เมื่อพิจารณาร่วมกับการหาความสัมพันธ์ระหว่างข้อมูลจากการประมวลผลภาษาธรรมชาติที่มีการวิเคราะห์หาความถี่ของชื่อถนนและข้อความที่เกี่ยวกับการเกิดอุบัติเหตุ และแผนที่ความหนาแน่นของข้อมูลระบุตำแหน่ง (geocoding) พบว่าความสอดคล้องกับการรายงานอุบัติเหตุและการจราจรจากหน่วยงานของภาครัฐซึ่งสอดคล้องกับผลงานวิจัยที่สามารถนำเหมืองข้อความ มาประยุกต์ใช้กับการทำเหมืองข้อมูลบนเว็ป (Web Mining) (วงศ์วิไลสกุล, 2556)

2.2 การประมวลผลภาษาธรรมชาติและการตัดคำด้วย PythaiNLP

การศึกษาการตัดคำเพื่อให้ได้คำที่ถูกต้องและมีความสมบูรณ์ที่สุด เนื่องจากชุดข้อมูลที่นำมาใช้ในงานวิจัยคือข้อมูลในรูปแบบข้อความที่บันทึกจากสื่อสังคมออนไลน์ที่เป็นภาษาไทย จึงเป็นเรื่องที่ท้าทายอย่างมากในการหาเทคนิคที่มีประสิทธิภาพในการจัดการกับข้อมูล เนื่องจากภาษาไทย มีการแบ่งคำและวรรณยุกต์ที่ยากและซับซ้อนกว่าภาษาอังกฤษ จากการทดลองและศึกษาพบว่า ข้อความภาษาไทยจะมีเครื่องมือการตัดคำที่ชื่อว่า PythaiNLP และใช้ไลบรารี newmm เป็นการตัดคำในขั้นตอนแรก และใช้ whitespace+newline ในการช่วยเว้นวรรคคำของ

ข้อความภาษาไทยให้มีความถูกต้องมากขึ้น จากการประมวลผลจะได้เป็นชื่อถนนและรายละเอียดของอุบัติเหตุจากข้อความที่ได้จากสื่อสังคมออนไลน์ (พิเดช, 2563)

2.3 การหาความสัมพันธ์ของข้อความด้วยอัลกอริทึมอพริโอริ

การประยุกต์ใช้กฎความสัมพันธ์เพื่อวิเคราะห์ความเสี่ยงของอุบัติเหตุจากข้อความด้วยอัลกอริทึม Apriori ในการหาความสัมพันธ์ของชื่อถนนและลักษณะการเกิดอุบัติเหตุ โดยข้อมูลที่นำมาสร้างกฎความสัมพันธ์รวบรวมมาจากสื่อสังคมออนไลน์ ซึ่งประกอบด้วยถนนที่เกิดอุบัติเหตุและข้อความรายละเอียดของเหตุการณ์อุบัติเหตุ มีความสอดคล้องกับการศึกษาการวิเคราะห์ความเสี่ยงการออกกลางคันของนักศึกษาที่มีค่าความเชื่อมันของกฎมากกว่าร้อยละ 90 (ศรีอุไร, 2561) แต่คำในภาษาไทยต้องมีการกำหนดค่าความเชื่อมั้นที่ต่ำกว่าคำในรูปแบบภาษาอังกฤษ เนื่องจากภาษาไทยมีความซับซ้อนของสระและวรรณยุกต์และการแบ่งเว้นช่องว่างที่แตกต่างกัน

3 ข้อเสนอแนะ

ชุดข้อมูลอุบัติเหตุจากสื่อสังคมออนไลน์ การวิเคราะห์เหมืองข้อความจากสื่อสังคมออนไลน์ สามารถวิเคราะห์ร่วมกับรูปภาพ ตัวอักษร ที่มีการแบ่งปันผ่านสื่อสังคมออนไลน์หลากหลายแพลทฟอร์ม ที่มีการระบุตำแหน่ง เช่น Foursquare, Flickr, Instagram โดยการเลือกใช้อัลกอริทึมและเครื่องมือช่วยเหลือในการวิเคราะห์รูปแบบข้อความอื่น ๆ ได้แก่ Thai tokenizer, ThaiNLP เพื่อศึกษารูปแบบการตัดแบ่งข้อความ เพื่อเป็นการพัฒนาการประมวลผลภาษาธรรมชาติในภาษาไทย ให้มีความแม่นยำเพิ่มมากขึ้นในอนาคต ข้อจำกัดของการรายงานอุบัติเหตุผ่านสื่อสังคมออนไลน์ในปัจจุบันยังมีการกระจุกตัวอยู่ในเมืองหลวง หากมีการแจ้งอุบัติเหตุผ่านระบบสื่อสังคมออนไลน์ที่เป็นแพลทฟอร์มสำหรับการแจ้งเหตุฉุกเฉินต่างๆ ที่นอกเหนือจากอุบัติเหตุจากการจราจรก็จะเป็นผลดีต่อประชาชนในการรับรู้ข่าวสารที่รวดเร็วมากยิ่งขึ้น อีกทั้งยังสามารถนำเทคโนโลยีที่น่าสนใจ การนำวิธีการเรียนรู้เชิงลึก 2 วิธี ได้แก่ Deep Belief Network (DNB) และ Long Shot Term Memory (LSTM) มาใช้ในการตรวจจับการจราจรหรืออุบัติเหตุที่ถูกกล่าวถึงได้

เครื่องมือการประมวลผลภาษาธรรมชาติจากที่ได้กล่าวข้างต้นในเรื่องของชุดข้อมูลที่มีลักษณะเฉพาะตัว งานวิจัยนี้เห็นถึงความสำคัญของเทคโนโลยีทางด้านไอทีที่นำมาใช้เพื่อการแก้ปัญหาต่าง ๆ ในโลกปัจจุบัน การพัฒนาเครื่องมือสำหรับการใช้แก้ปัญหาต่างๆเป็นสิ่งสำคัญอย่างมาก PyThaiNLP คือเครื่องมือที่นำมาใช้งานและเห็นว่ามีประโยชน์อย่างมากกับการใช้เป็น

เครื่องมือการประมวลผลสำหรับภาษาไทย ด้วยภาษาไทยเป็นภาษาที่มีความเป็นเอกลักษณ์เฉพาะตัว การพัฒนาให้เครื่องมือนี้มีประสิทธิภาพมากยิ่งขึ้นจะทำให้ผู้ที่ทำงานวิจัยเกี่ยวกับเรื่องประมวลผลภาษาไทยในอนาคตสามารถทำให้ประสิทธิภาพดีมากขึ้นและผลลัพธ์ที่ดีขึ้น

งานวิจัยในอนาคต การประมวลผลภาษาธรรมชาติจากข้อมูลสื่อสังคมออนไลน์เชิงพื้นที่เพื่อคาดการณ์อุบัติเหตุและเครื่องมือการประมวลผลภาษาธรรมชาติที่ถูกพัฒนาให้มีประสิทธิภาพดีขึ้นอย่างต่อเนื่องหากนำระบบการประมวลผลภาษาธรรมชาติจากข้อมูลสื่อสังคมออนไลน์มาประยุกต์ใช้กับระบบ Traffy Fondue ที่เป็นแพลตฟอร์มรับเรื่องและบริหารจัดการปัญหาเมือง ที่ประชาชนทุกคนสามารถแจ้งปัญหาที่พบเจอ อีกทั้งยังสามารถทำให้ระบบมีการแสดงผลแบบเรียลไทม์ ที่สามารถรับรู้ข้อมูลได้อย่างทันที เพื่อให้หน่วยงานต่าง ๆ ดำเนินการแก้ปัญหาเหล่านั้นอย่างรวดเร็ว โดยสามารถช่วยลดปัญหาการจราจรติดขัดที่เกิดจากอุบัติเหตุ และยังสามารถช่วยลดอุบัติเหตุจากการจราจรที่จะเกิดซ้ำซ้อน และพัฒนาระบบการประมาณค่าเชิงตำแหน่งของจุดการเกิดอุบัติเหตุที่มีความแม่นยำมากขึ้นเพิ่มประสิทธิภาพในการจัดการระบบคมนาคม ให้มีความปลอดภัยบนท้องถนนเพิ่มมากขึ้นในอนาคต

บรรณานุกรม

Aghajani, M. A., Dezfoulian, R. S., Arjroody, A. R., & Rezaei, M. (2017). Applying GIS to Identify the Spatial and Temporal Patterns of Road Accidents Using Spatial Statistics (case study: Ilam Province, Iran). *Transportation Research Procedia,* 25, 2126-2138. https://www.sciencedirect.com/science/article/pii/S2352146517307160

Agrawal, R., Imielinski, T. and Swami, A. (1993). Mining Association Rules between Sets of Items in Large Databases. Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data.

ANIKÓ KÁLMÁN, L. C. (2019). Complexity as new normality: What is going on? . *SEFI* 2019.

Antoine Limasset, C. M., Pierre Peterlongo, Lucie Bittner. (2016). Minimal perfect hash functions

in large scale bioinformatics Problem. *JOBIM* 2016.

Chakravarty, S. (2018). What is geocoding; how can it help sell products. https://www.geospatialworld.net/blogs/what-is-geocoding-and-how-can-it-help-sell-products/,2020

ESRI. (2021). What is geocoding?

H, M. (2020). การท า Machine Learning ด้วย Clustering Model.

Inc., F. (2020). *Facebook Inc. Certificate of Incorporation*. https://s21.q4cdn.com/399680738/files/doc_downloads/governance_documents/FB_CertificateOfIncorporation.pdf

Institute, S. (2020). การประมวลผลภาษาธรรมชาติ (Natural language processing). สืบค้นจาก https://www.sas.com/th_th/insights/analytics/what-is-natural-language-processing-nlp.html

Kanyawee Pornsawangdee, N. S. (2021). Association Rule: การหา Frequent Itemsets ด้วย Apriori Algorithm. https://bigdata.go.th/big-data-101/frequent-itemsets-apriori-algorithm/

Karagoz, P., Oguztuzun, H, Cakici, R, Ozdikis, O, Onal, K D and Sagcan, M. (2016). *Extracting Location Information from Crowd-sourced Social Network Data*. London: Ubiquity Press.

Karimi, H., Durcik, M., & Rasdorf, W. (2004). Evaluation of Uncertainties Associated with Geocoding Techniques. *Computer-Aided Civil and Infrastructure Engineering,* 19, 170-185. https://www.researchgate.net/publication/229889680_Evaluation_of_Uncertainties_Associated_with_Geocoding_Techniques

Karimi, H., Sharker, M., & Asavasuthirakul, D. (2011). Geocoding Recommender: An Algorithm to Recommend Optimal Online Geocoding Services for Applications. *T. GIS,* 15, 869-886. https://www.researchgate.net/publication/220606074_Geocoding_Recommender_An_Algorithm_to_Recommend_Optimal_Online_Geocoding_Services_for_Applications

Matci, D. K. (2018). *Computers, Environment and Urban Systems.* Turkey. (Research Institute of Earth and Space Science).

ncm, C. (2018). Python Opensource GIS GEOCODE. https://medium.com/geo-datascience/python-opensource-gis-07-f5c72a02b1c7

S., M. p., Savyan. (2023). An approach to detect compromised Twitter accounts using Cognitive Cloud Model and Multi-Attribute Decision Making.

Shekhar, H., Setty, S., & Mudenagudi, U. (2016). Vehicular traffic analysis from social media data. 2016 *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 1628-1634. https://ieeexplore.ieee.org/document/7732281

Steven Bird, E. K., Edward Loper. (2019). *Natural Language Processing with Python*.

Surapong, K. (2019). Geocode คืออะไร สืบค้นจาก https://www.bualabs.com/archives/2762/what-is-geocode-what-is-spatial-join-geodatagrame-difference-table-join-attribute-join-geospatial-ep-5/

Tanjil Ahmed, S. R., Niloy Routh,Eftakhar Alam Nirob. (2019). Market sales prospecting by

analyzing customer buying pattern

using machine learning.

Times, T. N. Y. (2010). "Why Twitter's C.E.O. Demoted Himself".
http://www.nytimes.com/2010/10/31/technology/31ev.html

TomTom. (2022). Bangkok traffic. https://www.tomtom.com/traffic-index/bangkok-traffic/

กลิ่นดาว, ส. (2542). ระบบสารสนเทศภูมิศาสตร์ : หลักการเบื้องต้น. กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยธรรมศาสตร์.

คะอังกุ, น. (2555). วิธีแบ่งช่วงข้อมูลสำหรับการหากฎความสัมพันธ์.

คำภีรานนท์, พ. (2558). เทคโนโลยีการจราจร. In.

คีรีทอง, อ. (2555). เทคโนโลยีทางภูมิศาสตร์.

งานป้องกันและบรรเทาสาธารณภัย. (2553). สาเหตุของอุบัติเหตุจราจรทางบกและหลักสำคัญ.

จิรขจรกุล, ส. (2555). เรียนรู้ระบบสารสนเทศภูมิศาสตร์ด้วยโปรแกรม *ArcGIS* 10.1 *for Desktop*. ปทุมธานี: คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต.

ชิษณุ, ว., & ชัยพล. (2550). การพิสูจน์จุดอันตรายต่อการเกิดอุบัติเหตุบริเวณถนนรามอินทราและถนนแจ้งวัฒนะ. *NCCE,* ครั้งที่ 12, 189-201.

ทองสว่าง, ท. (2520). ระบบสารสนเทศภูมิศาสตร์. กรุงเทพมหานคร : สำนักพิมพ์โอเดียนสโตร์

นงนุช, ก., ซาเฮาะ, อ., & ตุ้มทอง, ส. (2557). การวิเคราะห์บทความอัตโนมัติโดยใช้กระบวนการภาษาธรรมชาติ. การประชุมวิชาการระดับชาติมหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ, ครั้งที่ 1, 472-479.

บัวทอง, ว. (2557). Association Rules.

พิเดช, เ. (2563). ระบบจำแนกตั๋วงานอัจฉริยะด้วยเทคนิควิเคราะห์ข้อความไทย-อังกฤษ และกำรเรียนรู้ของเครื่อง.

พิศิษฐ์ บวรเลิศสุธี, ว. ไ. (2565). ตัวแบบการวิเคราะห์ความรู้สึกทางอารมณ์สำหรับจำแนกประเภทบทความแนะนำสินค้าออนไลน์. *Journal of Engineering and Digital Technology*.

ลววิโรจน์วงศ์, ส. (2557). ฐานข้อมูลสารสนเทศภูมิศาสตร์.

วงศ์วิไลสกุล, ว. (2556). เหมืองข้อความและการประยุกต์ใช้.

วิกาหะ, ป. (2557). การศึกษาเกี่ยวกับการรู้จำชื่อเฉพาะ. *KMITL Information Technology Journal*.

วิจิตรบุญยรักษ์, พ. (2553). สื่อสังคมออนไลน์: สื่อแห่งอนาคต. *Excutive Journal*, 99-103.

ศรีอุไร, ป. ป. แ. (2561). การประยุกต์ใช้กฎความสัมพันธ์เพื่อวิเคราะห์ความเสี่ยงการออกกลางคันของนักศึกษาสาขาเทคโนโลยีสารสนเทศ. วารสารวิทยาศาสตร์และวิทยาศาสตร์ศึกษา.

ศูนย์วิจัยภูมิสารสนเทศเพื่อประเทศไทย. ระบบสารสนเทศภูมิศาสตร์. http://www.gisthai.org/about-gis/gis.html

สำนักงานเทคโนโลยีอวกาศและภูมิศาสตร์. (2552). ระบบสารสนเทศภูมิศาสตร์ : หลักการเบื้องต้น.

สำนักงานจราจรและการขนส่ง. (2562). การวิเคราะห์ข้อมูลสารสนเทศภูมิศาสตร์.

สำนักงานปลัดกระทรวงคมนาคม, ศ. (2564). อุบัติเหตุบนโครงข่ายถนนของกระทรวงคมนาคม. https://datagov.mot.go.th/dataset/roadaccident

สำนักปลัดกรุงเทพมหานคร. (2550). กรุงเทพมหานคร.

สิทธิ, อ. (2561). การวิเคราะห์เหมืองดัชนีถ้อยคำจากข้อมูลระบุตำแหน่งเชิงพื้นที่ผ่านสื่อสังคมออนไลน์. วารสารสังคมศาสตร์ มศว, 304-319.

อยู่เมือง, ส. (2540). การประยุกต์ใช้ระบบสารสนเทศทางภูมิศาสตร์เพื่อการบริหาร.

อรวรรณ วงศ์แก้วโพธิ์ทอง. (2553). Social Media เครื่องมือเพิ่มศักยภาพทางธุรกิจ. *Excutive Journal*, 63-69.

อุทัยเศรษฐวัฒน์, ท. (2550). การประยุกต์ใช้ระบบสารสนเทศภูมิศาสตร์ในการหาจุดอันตรายบนทางพิเศษ. มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพฯ.

ประวัติผู้เขียน

ชื่อ-สกุล	สิทธิพัฒน์ เปี่ยมใจสว่าง
วัน เดือน ปี เกิด	13 มิถุนายน 2540
สถานที่เกิด	จังหวัดสมุทรปราการ
วุฒิการศึกษา	พ.ศ. 2561 วิทยาศาสตรบัณฑิต สาขาวิชาภูมิศาสตร์และภูมิสารสนเทศ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพมหานคร