การวิเคราะห์เหมืองดัชนีถ้อยคำ จากข้อมูลระบุตำแหน่งเชิงพื้นที่ผ่านสื่อสังคมออนไลน์
Tags Mining Analysis Using Geotagged Online Social Media Data

อสมาภรณ์ สิทธิ[1]
Asamaporn Sitthi

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์เหมืองดัชนีถ้อยคำจากข้อมูลระบุตำแหน่งเชิงพื้นที่ผ่านสื่อสังคมออนไลน์ แอปพลิเคชันฟลิกค์เกอร์ (flickr) ในประเทศไทย ข้อมูลที่ได้จากการระบุตำแหน่ง (geo-tagged) มาจากการที่ผู้ใช้งานแบ่งปันข้อมูล รูปภาพหรือแสดงความคิดเห็นต่างๆ ที่มีความสัมพันธ์กับสถานที่ งานวิจัยนี้นำเทคนิคเหมืองดัชนีข้อความ (tags mining) โดยการนำข้อมูลผ่านกระบวนการสกัดความรู้ มาวิเคราะห์เพื่อค้นหารูปแบบหรือความสัมพันธ์ของข้อมูล การดำเนินการวิจัยแบ่งออกเป็น 3 ขั้นตอนหลักคือ 1) ดึงข้อมูลระบุตำแหน่งและสร้างคลังฐานข้อมูล 2) ประมวลผลข้อความ 3) จัดกลุ่มแบบลำดับขั้นและหากฎของความสัมพันธ์ระหว่างข้อความและวิเคราะห์ความหนาแน่นเชิงพื้นที่ ผลการวิจัยที่ได้จากการจัดกลุ่มและการหากฎของความสัมพันธ์ดัชนีข้อความและการวิเคราะห์ความหนาแน่นที่อธิบายถึงลักษณะการกระจายเชิงพื้นที่ของการระบุตำแหน่ง flickr สรุปได้ว่า ดัชนีถ้อยคำ flickr ที่ความคล้ายคลึงกันมากในแต่ละกลุ่มจะมีค่าความเชื่อมั่นสูง ซึ่งสามารถแบ่งกลุ่มของข้อความออกเป็น 3 ประเภทที่สำคัญ ได้แก่ ประเภทสถานที่ท่องเที่ยวเชิงธรรมชาติ สถานที่ท่องเที่ยวเชิงวัฒนธรรม และกิจกรรมพิเศษอื่นๆ ซึ่งเชื่อมโยงกับลักษณะการใช้งานสื่อสังคมออนไลน์ และแสดงให้เห็นถึงปัจจัยต่างๆ ที่มีผลต่อพฤติกรรมและการตอบสนองของผู้ใช้งาน ได้แก่ ความนิยมของสถานที่ท่องเที่ยว ตามลักษณะภูมิประเทศและฤดูกาล ซึ่งสามารถนำไปใช้ประโยชน์ในการวางแผนและพัฒนาในด้านต่างๆ การเข้าถึงพื้นที่ โครงสร้างพื้นฐาน การให้บริการเครือข่ายอินเทอร์เน็ต คมนาคมขนส่ง และกิจกรรมต่างๆ ที่มีความสำคัญต่อสถานที่นั้นๆ

คำสำคัญ: *เหมืองดัชนีข้อความ การระบุตำแหน่ง ภูมิสารสนเทศศาสตร์*

---

[1]อาจารย์ประจำภาควิชาภูมิศาสตร์ คณะสังคมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ
Lecturer, Department of Geography, Faculty of Social Sciences, Srinakarinwirot University, Bangkok.
Coresponding e-mail: asamaporn@g.swu.ac.th

Abstract

This research aims to use the tag mining techniques to analyze the Geotagged Flicker data. The Flickr plays an important role in the online social media platform, and has been intensively used by a number of users to share text and photos from different part of the world. Tags mining algorithms is implemented to extract knowledge and analyze tags for finding the clustered and its associate. The research methodology consists of 1) querying data via Flickr API and perform geotagged corpus database 2) pre-processing and filtering the tags 3) mining tags using hierarchical cluster analysis, associate rule technique, and mapping density of Flickr distribution at spatio-temporal scale. The results show that flickr tags can be clustered into 3 main groups including natural tourist attraction, cultural tourist attraction and special activities and others. Moreover, the association rule result is also confirmed that these 3 main groups of flickr's tags has high correlation from the confidence values. The distribution of Flickr data can capture the user's behavior and responses related to particular locations. This information can be an indicator of popularity, topography, seasons, accessibility, transportation and special activities associated with those locations.

Keywords: *Tags mining, Geotagged, Geoinformatic*

บทนำ

ในยุคปัจจุบันการพัฒนาทางเทคโนโลยีสารสนเทศเพื่อใช้ในการติดต่อสื่อสารระหว่างกันเป็นสิ่งสำคัญ เครือข่ายให้บริการทางอินเทอร์เน็ตทำให้การสื่อสารสามารถเข้าถึงได้ทุกคนทุกที่ทุก ในการติดต่อเชื่อมโยง ผู้ใช้งานทั่วทุกมุมโลกสามารถติดต่อ แลกเปลี่ยนและแบ่งปันความคิดเห็น ข้อมูลเหตุการณ์หรือกิจกรรมที่สนใจในหลากหลายมุมมอง ผ่านตัวอักษร เสียง รูปภาพ และวิดีโอ นอกจากนี้ผู้ใช้งานยังสามารถแบ่งปันข้อมูลจากการบริการการบอกตำแหน่งทางภูมิศาสตร์ (Location Based Service: LBS) ที่มีการระบุสถานที่หรือตำแหน่งของสถานที่นั้นๆ โดยใช้อุปกรณ์พกพา เช่น โทรศัพท์เคลื่อนที่ ในแต่ละวันมีจำนวนผู้ใช้สื่อสังคมนี้เพิ่มมากขึ้นจนเรียกได้ว่าเข้ามาเป็นส่วนหนึ่งในชีวิตประจำวัน จำนวนข้อมูลและความถี่ที่มีการระบุตัวตนของสถานที่หรือตำแหน่งนั้นๆ ไม่ว่าจะเป็นจากผู้มาเยือนหรือนักท่องเที่ยวล้วนแต่มีความสำคัญในการติดตามหรือบ่งชี้ถึงความนิยมและสิ่งอำนวยความสะดวกต่างๆ ทั้งยังช่วยในเรื่องของการวางแผนและการจัดการทรัพยากรในสถานที่นั้นอีกด้วย การพัฒนาทางเทคโนโลยีสารสนเทศเพื่อใช้ในการติดต่อสื่อสารระหว่างกันเป็นสิ่งที่จำเป็นและเข้ามามีบทบาทสำคัญสำหรับคนในยุคปัจจุบัน

ข้อมูลที่ได้จากสื่อสังคมออนไลน์เหล่านี้มีลักษณะในการใช้งานผ่านทางแอปพลิเคชั่นต่างๆ ตามวัตถุประสงค์และความนิยมในการแบ่งปันข้อมูล เช่น เฟสบุ๊ค (facebook) ฟลิกเกอร์ (flickr) ทวิตเตอร์ (twitter) อินสตาแกรม (instragram) โฟร์สแควร์ (foursqaure) เป็นต้น แพลทฟอร์มที่ได้รับความนิยมและใช้กันอย่างแพร่หลายมากที่สุดได้แก่ facebook หากแต่การเข้าถึงข้อมูลของนักพัฒนายังมีจำกัด flickr เป็นหนึ่งในแพลตฟอร์มสื่อสังคมออนไลน์ที่ถูกพัฒนาขึ้นก่อนแพลทฟอร์มอื่นๆ โดยมีเอกลักษณ์ในการแบ่งปันภาพถ่ายรุบุตำแหน่งในสถานที่ต่างๆ และถูกนำมาศึกษาและใช้งาน

มากขึ้นในด้านวิทยาศาสตร์สิ่งแวดล้อม (Huang & Gartner. 2018) นอกจากนี้ twitter เป็นแพลทฟอร์มที่ได้รับความนิยมที่มีการแลกเปลี่ยนข้อมูลสนทนาที่มีลักษณะกระชับและมีการกระจายข้อมูล (retweet) ได้อย่างรวดเร็วร่วมกับการตอบกลับความคิดเห็น (reply) ในส่วนของ instagram เป็นแพลทฟอร์มใหม่ที่เพิ่งก่อตั้งในปี 2510 มีลักษณะในการแลกเปลี่ยนข้อมูลทั้งรูปภาพ วิดีโอ และเรื่องราว (story) หากแต่การเข้าถึงข้อมูลเชิงตำแหน่งสาธารณะยังถูกจำกัด จะเห็นได้ว่าในแต่ละแพลตฟอร์มมีลักษณะข้อดีและข้อจำกัดขึ้นอยู่กับวัตถุประสงค์ของผู้ใช้นั้นๆ ด้วยเหตุที่กล่าวมาข้างต้นเป็นจึงมีการนำข้อมูลสื่อสังคมออนไลน์มาใช้ประโยชน์ในด้านต่างๆ เพื่อตอบสนองต่อความต้องการและโอกาสใหม่ๆ ได้อย่างมีประสิทธิภาพและเป็นสาธารณะ (Richmond and Riva. 2010)

สำหรับระบบสารสนเทศภูมิศาสตร์ เป็นหนึ่งในเทคโนโลยีที่เข้ามามีบทบาทสำคัญในการจัดการเชิงพื้นที่ทั้งในการดำเนินชีวิต ภาครัฐและเอกชน อย่างไรก็ตาม ปัจจัยแวดล้อมหลายประการที่เป็นประเด็นที่น่าสนใจในการได้มาซึ่งข้อมูล การหาตำแหน่ง การแลกเปลี่ยนข้อมูล หรือผลกระทบที่เกิดจากการเปลี่ยนแปลงเชิงพื้นที่อย่างรวดเร็ว ซึ่งล้วนเป็นบริบทที่น่าสนใจในการเชื่อมโยงข้อมูลในมิติที่เหมาะสม (คณะกรรมการภูมิสารสนเทศแห่งชาติ. 2559) การใช้เทคโนโลยีจากสื่อสังคมออนไลน์ (social sensing technology) ทำให้ได้มาซึ่งข้อมูลที่มีขนาดใหญ่ (big data) ที่ได้จากโทรศัพท์มือถือทำงานร่วมกับอุปกรณ์ในการระบุตำแหน่งบนโลก (GPS) ซึ่งมีความแตกต่างกันจากแหล่งข้อมูลแบบดั้งเดิมในการนำเทคโนโลยีสื่อสารมาใช้ในการได้มาซึ่งข้อมูล และมีความน่าสนใจในแง่มุมต่างๆ ทางภูมิศาสตร์ ซึ่งข้อมูลเหล่านี้จะมีการฝังข้อมูลของข้อมูล (Metadata) ที่ได้จากการแบ่งปันข้อมูลออนไลน์สาธารณะนั้น ประกอบไปด้วย ข้อความที่เกี่ยวข้องกับเหตุการณ์หรือกิจกรรมการแบ่งปัน วันที่และเวลาทำการแบ่งปัน ดัชนีถ้อยคำ (hashtags หรือ tags) ที่เกี่ยวข้องสถานการณ์นั้น และข้อมูลตำแหน่งที่ตั้งของสถานที่หรือผู้ใช้งานจากอุปกรณ์รับสัญญาณจากดาวเทียม (GPS) ซึ่งจะเรียกการรวมกันของข้อมูลนี้ว่าการสะกดรอยตามข้อมูลที่ระบุพิกัด (geo - tagged) จะเห็นได้ว่า การสะกดรอยตามข้อมูลที่ระบุพิกัดจำนวนมากเหล่านี้สามารถสะท้อนให้เห็นถึงความนิยมและพฤติกรรมการใช้งานในรูปแบบต่างๆ ของผู้ใช้ซึ่งมีความสัมพันธ์เชิงพื้นที่ได้ เช่น การท่องเที่ยว ติดตามข่าวสาร การคมนาคมขนส่ง การเตือนภัยภัยพิบัติ การทำธุรกิจ การศึกษา การตรวจสอบความถูกต้องในพื้นที่จริง รวมไปถึงเศรษฐกิจ สังคม และวัฒนธรรม เป็นต้น ด้วยวิวัฒนาการการใช้เทคโนโลยีในการสื่อสารเป็นไปอย่างรวดเร็ว จำนวนข้อมูลที่เกิดขึ้นในแต่ละวันนั้นมีมากมายมหาศาล จึงเกิดเป็นข้อมูลที่เรียกว่า ฝูงข้อมูล (crowdsourcing) หรือ แผนที่บนเว็บไซต์ที่สามารถแสดงได้โดยไม่มีค่าใช้จ่าย ซึ่งข้อมูลเหล่านี้เกิดจากอาสาสมัครในระบบภูมิสารสนเทศ (Volunteer Geographic Information System: VGI)

จากข้อมูลข้างต้นนั้น เพื่อให้การค้นหาความหมายหรือนัยในการใช้งาน เหมืองข้อมูล (data mining) เข้ามามีบทบาทในการเพิ่มประสิทธิภาพของการวิเคราะห์ข้อมูล เพื่อค้นหารูปแบบ แนวทางและความสัมพันธ์ที่แฝงอยู่ในชุดข้อมูลเหล่านี้ โดยอาศัยหลักสถิติและการรู้จำในการประมวลผล ผู้วิจัยเล็งเห็นถึงประโยชน์ในการใช้ข้อมูลการรับรู้สื่อสังคมสำหรับภูมิสารสนเทศ จำนวนมหาศาลในรูปแบบใหม่ กรณีศึกษาแพลตฟอร์ม flickr มาวิเคราะห์เหมืองดัชนีถ้อยคำ เพื่อให้ทราบถึง แนวคิดเกี่ยวกับการรับรู้ทางสังคม การได้มาซึ่งข้อมูล รูปแบบพฤติกรรมในการใช้สื่อสังคมและการตอบสนองของผู้ใช้งาน flickr แนวโน้มการใช้งานสื่อสังคมออนไลน์ที่มีส่วนในการช่วยพัฒนาองค์ความรู้และการประยุกต์ทางด้านภูมิสารสนเทศ โดยการวิเคราะห์รูปแบบการแสดงสถานะผ่านตัวอักษรที่สกัดจากดัชนีถ้อยคำและการวิเคราะห์เหมืองข้อมูลจากการระบุตำแหน่งสามารถสื่อได้ถึงสิ่งที่มีความเกี่ยวข้องกับสถานที่นั้นๆ ซึ่งจะเป็นประโยชน์สำหรับการใช้งานในอนาคตได้อย่างเหมาะสม

วัตถุประสงค์

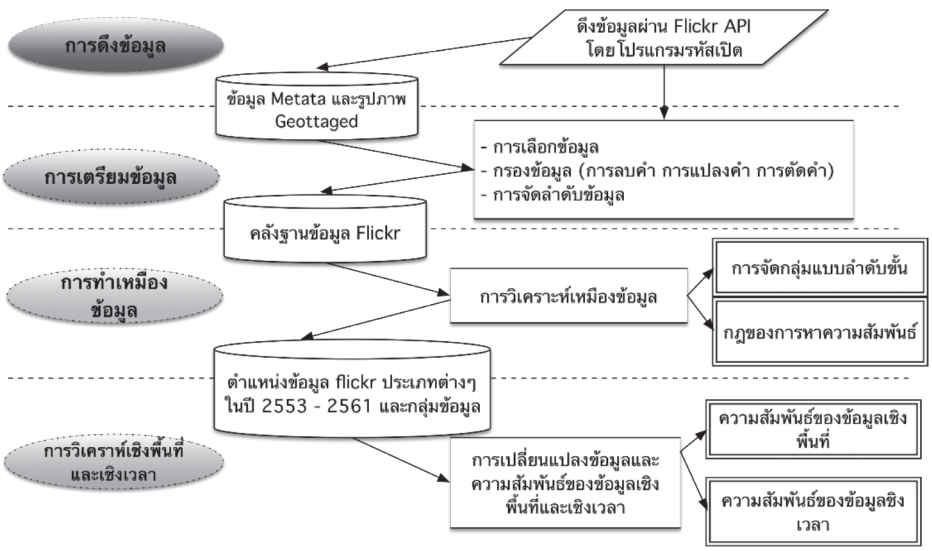
1. สกัดดัชนีถ้อยคำ flickr และพัฒนาคลังฐานข้อมูลเชิงพื้นที่
2. วิเคราะห์เหมืองดัชนีถ้อยคำเพื่อหาความสัมพันธ์ระหว่างข้อความและข้อมูลเชิงตำแหน่ง

ขอบเขตของการวิจัย

งานวิจัยนี้สกัดดัชนีข้อความจากการแบ่งปันแลกเปลี่ยนข้อมูลภาพและสถานที่ในรูปแบบของการระบุตำแหน่งทางภูมิศาสตร์ (geo-tagged) ผ่านการเรียกใช้เอพีไอ (Application Programming Interface: API) ในด้านของแพลตฟอร์มเว็บโปรแกรมมิ่ง (web programming platform) ได้แก่ ฟลิกเกอร์ (flickr) และนำเสนอวิธีการคือวิธีในการสร้างเครื่องมือเพื่อพัฒนาคลังข้อมูลเชิงพื้นที่โดยรวบรวมข้อมูล กรณีศึกษา พื้นที่ประเทศไทยตั้งแต่ปี 2553 ถึง 2561 โดยจะนำข้อมูลจัดเก็บแบบไม่มีโครงสร้าง และไม่อยู่ในรูปแบบของข้อมูลทางภูมิศาสตร์ ไปสร้างฐานข้อมูลแบบมีโครงสร้าง มุ่งเน้นไปที่ข้อมูลที่สามารถเปลี่ยนแปลงได้ เช่น ข้อมูลดัชนีถ้อยคำ (tags) ข้อมูลระบุตำแหน่ง (geo-tagged) รวมไปถึงการวิเคราะห์รูปแบบการดำเนินชีวิตซึ่งสะท้อนให้เห็นถึงลักษณะการใช้สื่อสังคมออนไลน์ทั้งในเชิงพื้นที่และเชิงเวลา

วิธีดำเนินการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ลักษณะการใช้งานเชิงตำแหน่งจากการใช้งานสื่อสังคมออนไลน์โดยวิธีการทำเหมืองข้อมูลซึ่งเป็นกระบวนการที่ค้นหารูปแบบและความสัมพันธ์ของข้อมูล ซึ่งแบ่งตามลักษณะวิธีดำเนินงานวิจัยแบ่งออกเป็น 4 กระบวนการหลัก ได้แก่ การรวบรวมโดยการสืบค้นข้อมูลที่เหมาะสมทั้งเนื้อหาและปริมาณข้อมูลจำนวนมาก การสร้างคลังฐานข้อมูล การประมวลผลข้อความ และหาความสัมพันธ์ระหว่างข้อมูลทั้งในเชิงพื้นที่และเชิงเวลาจากการตอบสนองของผู้คนจากข้อมูลระบุตำแหน่งผ่านสื่อสังคมดังภาพที่ 1



ภาพที่ 1: วิธีดำเนินการวิจัย

1. การสืบค้นข้อมูลและการสร้างคลังฐานข้อมูล (corpus document)

งานวิจัยนี้รวมรวมข้อมูลโดยการสืบค้นข้อมูลจากเว็บเครือข่ายสังคม flickr ผ่านการเรียกใช้เอพีไอ (API) ของเว็บไซต์ flickr (www.flickr.com) (Sergieh et. al. 2012) โดยโปรแกรมอาร์ (R programming) เพื่อดึงข้อมูลสาธารณะจากผู้ใช้งานทั่วไปที่มีการอัปโหลดในสื่อสังคมออนไลน์ โดยมีการพัฒนาโปรแกรมเพื่อทำการดึงข้อมูล ในรูปแบบของ Representational state transfer หรือ REST ซึ่งเป็นการดึงข้อมูลผ่าน URL ในลักษณะของการร้องขอข้อมูลตามเงื่อนไขที่กำหนดเพื่อให้ได้รับข้อมูลส่งกลับมาอยู่ในรูปของ Extensible Markup Language หรือ XML ซึ่งเป็นรูปแบบเอกสารที่เป็นตัวอักษรและรองรับการเข้ารหัส (unicode) ได้ทุกภาษา โดยข้อมูลเหล่านี้จะต้องมีการระบุเงื่อนไขในการได้มาซึ่งข้อมูลของข้อมูล (metadata) ประกอบไปด้วย ชื่อผู้ใช้งาน (username) วันเละเวลาที่อัปโหลดข้อมูล (date and time) ดัชนีถ้อยคำ (tags) และการระบุตำแหน่งของข้อมูล (geo-tagged) และนำข้อมูลที่ไม่มีโครงสร้างในรูปแบบข้อความ (ภาพที่ 2) มาแปลงให้อยู่ในรูปแบบมีโครงสร้างที่แน่นอนเพื่อจัดทำเป็นคลังฐานข้อมูล ซึ่งเป็นข้อมูลที่มีความสำคัญในการนำไปวิเคราะห์ลักษณะการดำเนินชีวิตทั้งในเชิงพื้นที่และเชิงเวลาต่อไป

```
    <photo id="5141963968" owner="26515295@N06" secret="4035dea420" serv
er="1225" farm="2" title="OLYMPUS DIGITAL CAMERA" ispublic="1" isfriend=
"0" isfamily="0" datetaken="2010-10-18 19:03:04" datetakengranularity="0
" datetakenunknown="0" tags="travel geotagged thailand bangkok skybar st
atetower 10182010 geo:lat=137210133333339 geo:lon=100517201111111" latit
ude="13.721013" longitude="100.517201" accuracy="16" context="0" place_i
d="0CfrewNTWrnNlprWMA" woeid="28349065" geo_is_family="0" geo_is_friend=
"0" geo_is_contact="0" geo_is_public="1"/>
```

ภาพที่ 2: การได้มาซึ่งข้อมูล flickr ที่ไม่มีโครงสร้าง

2. การวิเคราะห์เหมืองดัชนีถ้อยคำ

ข้อมูลที่ดึงมาได้จาก flickr นั้นจะเป็นข้อความที่มีการบรรยายผ่านดัชนีถ้อยคำ (tags) ร่วมกับลักษณะการระบุตัวตน (geo-tagged) จากข้อมูลสื่อสังคมออนไลน์ ได้แก่ ข้อเท็จจริงหรือเหตุการณ์ต่าง ๆ ที่เกิดขึ้น ความรู้สึกนึกคิดหรืออารมณ์ที่ผู้ใช้แสดงความคิดเห็นผ่านตัวอักษร และข้อเสนอแนะ (กานดา แผ่วัฒนกุล และ ปราโมทย์ ลือนาม. 2556) อย่างไรก็ตาม ข้อมูลสื่อสังคมออนไลน์ flickr เป็นการใช้เพื่อตอบสนองความต้องการในการแบ่งปันรูปภาพต่อสาธารณะ ข้อความที่มีการบรรยายจะถ่ายทอดในลักษณะบ่งชี้หรืออธิบายถึงสิ่งที่อยู่ในรูปภาพ กล้องถ่ายรูป วันเวลา บุคคล กิจกรรมหรือลักษณะของสถานที่นั้นๆ ซึ่งสามารถแสดงให้เห็นถึงทัศนคติของผู้ใช้โดยแบ่งออกเป็นขั้วความคิดเห็นหรือความนิยมได้ การวิเคราะห์เหมืองดัชนีถ้อยคำนั้นจะแบ่งออกเป็น 2 ส่วนหลักๆ ได้แก่ การประมวลผลข้อความ (text processing) และการหากฎความสัมพันธ์ของเหมืองดัชนีถ้อยคำ (Association Rule Mining: ARM) โดยมีรายละเอียดดังนี้

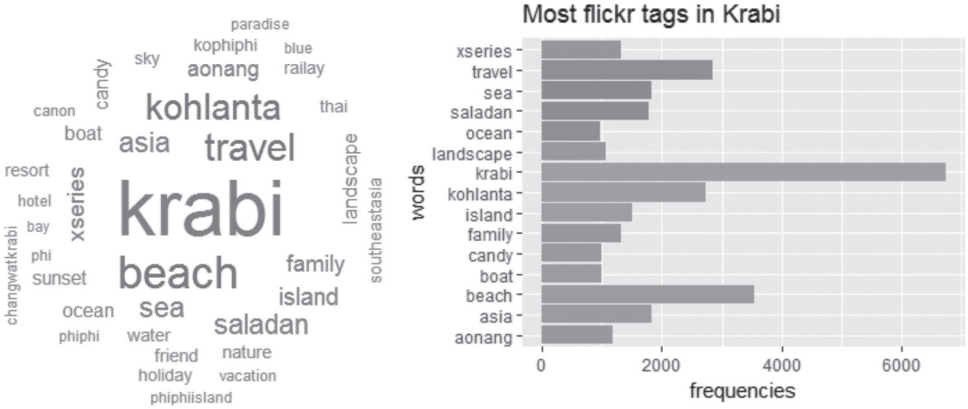
2.1 การประมวลผลข้อความ (Text processing)

การประมวลผลข้อความเป็นขั้นตอนการเตรียมข้อมูลสำหรับการวิเคราะห์เหมืองดัชนีถ้อยคำจากสื่อสังคมออนไลน์เพื่อให้ได้ข้อมูลที่เหมาะสมและบิดเบือนน้อยที่สุด ประกอบด้วย 2 กระบวนการ ได้แก่ การเลือกคุณลักษณะของข้อมูลและการกรองข้อมูล ซึ่งมีรายละเอียดดังนี้

2.1.1 การเลือกคุณลักษณะของข้อมูล การเลือกคุณลักษณะของข้อมูลเป็นการวิธีการเบื้องต้นในการคัดเลือกข้อมูลที่ไม่มีความจำเป็นเพื่อลดความคลาดเคลื่อนต่อการวิเคราะห์ เช่น การลบคำที่ไม่สำคัญ (stop words) ได้แก่ ตัวเลข ลักษณะนาม ข้อความที่เกี่ยวข้องกับอุปกรณ์การได้มาซึ่งข้อมูล หรือคำที่ไม่มีผลต่อการวิเคราะห์ และคำซ้ำต่างๆ การแปลงคำให้อยู่ในรูปแบบของรากศัพท์ (stemming) จากเทคนิคพจนานุกรม และการตัดคำออกมาเป็นคำเดี่ยวๆ (tokenization) เพื่อนำไปจัดกลุ่มและหากฎของความสัมพันธ์ต่อไป (เอกสิทธิ์ พัชรวงศ์ศักดา. 2557; Mahgoub et al. 2008)

2.1.2 การกรองข้อมูล การกรองข้อมูลแบ่งออกเป็น 2 ประเภท ได้แก่ การกรองข้อมูลที่เป็นข้อความและการกรองข้อมูลเชิงตำแหน่ง ซึ่งการกรองข้อมูลข้อความเป็นการกลั่นกรองและแก้ไขข้อมูลคำผิดหรือบิดเบือนไปจากปกติ (outlier data) เช่น ตัวสะกด โดยเฉพาะชื่อสถานที่หรือคำที่ระบุถึงตำแหน่งต่างๆ ให้ถูกต้อง ในส่วนของการกรองข้อมูลเชิงตำแหน่งนั้นอาจมีความคลาดเคลื่อนเนื่องมาจากอุปกรณ์ที่ใช้ในการระบุตำแหน่ง หรือมีการระบุตำแหน่งที่นอกเหนือจากขอบเขตที่ศึกษาจึงต้องมีการกรองข้อมูลเชิงตำแหน่งเฉพาะในขอบเขตการศึกษา

2.1.3 การจัดลำดับข้อมูล โดยวิเคราะห์จำนวนความถี่ของคำที่เกิดขึ้นในข้อความนั้นๆ ให้อยู่ในรูปแบบถุงของข้อมูล (bag of words) จัดลำดับข้อมูลและนำมาแสดงผลในรูปแบบของดัชนีถ้อยคำก้อนเมฆ (tagcloud) และกราฟแสดงความถี่ของดัชนีถ้อยคำ (frequency of words) ดังภาพที่ 3



ภาพที่ 3: ดัชนีถ้อยคำก้อนเมฆ (tagcloud)
และกราฟแสดงความถี่ของดัชนีถ้อยคำ (frequency of words)

2.2. การหาความสัมพันธ์ของเหมืองดัชนีถ้อยคำ

2.2.1 การจัดกลุ่มข้อมูล (clustering) การจัดกลุ่มข้อมูลจะพิจารณาจากลักษณะการเกิดขึ้นของแต่ละชุดข้อมูล flickr ตามคุณลักษณะ (feature) เช่น คน สัตว์ สิ่งของ สถานที่ ภูมิทัศน์ กิจกรรมต่างๆ หรือ ขั้วความคิดเห็น (polarity) (ข้อเสนอแนะและความรู้สึก) ฯลฯ ตามช่วงเวลาการแบ่งปันข้อมูลในแต่ละสถานที่ เพื่อประโยชน์ในลดมิติของข้อมูล โดยในการจัดกลุ่มประเภทของข้อมูลนั้นจะแบ่งออกเป็นคลัสเตอร์ (cluster) จากความคล้ายคลึงกันของข้อมูล โดยดัชนีถ้อยคำ (tags) ทั้งหมดที่สกัดได้จะถูกให้คะแนนและค่าน้ำหนักตามความถี่ของดัชนีถ้อยคำ (term of frequency) แล้วจะถูกแปลงให้อยู่ในรูปเอกสารเมทริกต์ (terms document matrix) เพื่อจัดระเบียบ ลดมิติและขนาดของข้อมูลที่ซับซ้อน โดยแปลงเอกสารเมทริกต์จะถูกแปลงเป็นสเกลอันตรภาค (Interval Scale) เพื่อจัดทำข้อมูลให้เป็นมาตรฐานก่อน หลังจากนั้นจะนำเอกสารเมทริกต์มาจัดกลุ่มตามน้ำหนักของข้อมูลที่มีความคล้ายคลึงกัน

งานวิจัยนี้ใช้เทคนิค การจัดกลุ่มเป็นลำดับขั้น (Hierarchical Cluster Analysis) ซึ่งเป็นเทคนิคที่นิยมใช้กับข้อมูลเชิงคุณภาพและข้อมูลเชิงปริมาณ เป็นการจัดกลุ่มแบบไม่กำกับดูแล (unsupervised) โดยไม่จำเป็นต้องทราบประเภทและจำนวนกลุ่ม (Jomsri. 2012) การจัดกลุ่มดัชนีถ้อยคำโดยเทคนิคลำดับขั้นนั้นจะวัดค่าความคล้ายคลึง โดยพิจารณาความแตกต่างหรือความผันแปรที่น้อยที่สุด ตามวิธีของ Ward's Minimum Variance method และเปรียบเทียบค่าเฉลี่ยของระยะทางภายในกลุ่มและระหว่างกลุ่มดัชนีถ้อยคำที่จัดได้ (Murtagh. 2009) และแสดงผลโดยใช้แผนภาพแสดงการจัดกลุ่ม (dendrogram) เพื่อศึกษาความสัมพันธ์ระหว่างดัชนีถ้อยคำ

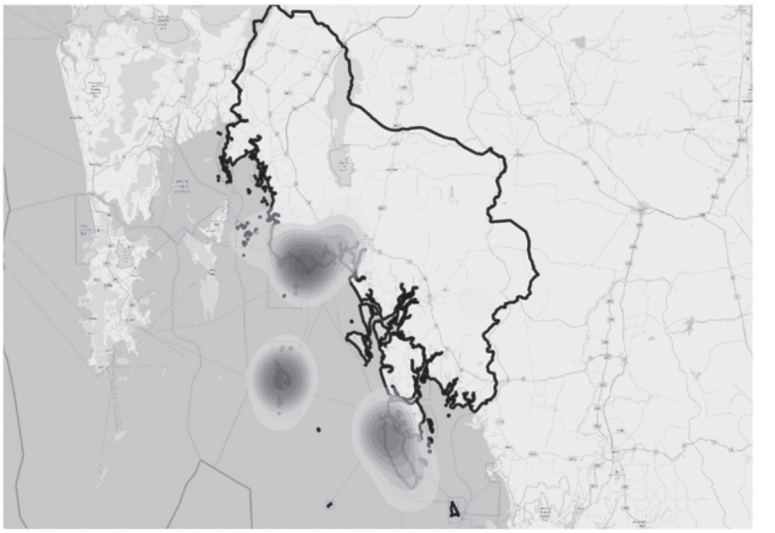
2.2.2 การหาความสัมพันธ์ระหว่างข้อมูล

1) กฎของการหาความสัมพันธ์ (Association Rule Mining: ARM) ขั้นตอนการหาความสัมพันธ์ระหว่างคำนี้ประยุกต์กฎของการหาความสัมพันธ์ของข้อมูลจากดัชนีถ้อยคำจะแยกกฎของความสัมพันธ์ออกจากกันโดยอัตโนมัติ โดยกำหนดชุดของดัชนีถ้อยคำสำคัญหลัก (K) 10 คำ จากข้อมูลที่ผ่านการกรองข้อมูลความถี่เชิงพื้นที่ (spatial frequency) 10 จังหวัดที่มีการใช้งาน flickr มากที่สุด มาจัดทำดัชนีคอลเล็กชันของดัชนีถ้อยคำ Ki = {k1, k2, ... , kn} เมื่อ n คือจำนวนประโยคทั้งหมดในเอกสาร และหาความสัมพันธ์ระหว่างดัชนีถ้อยคำต่างๆ (ki) ที่เกี่ยวข้องกับดัชนีถ้อยคำสำคัญหลัก (K) ในกรณีที่ ki มีความสัมพันธ์กับ K ($ki \subseteq K$) มากนั้น กล่าวได้ว่าดัชนีถ้อยคำ ki มักจะพบอยู่กับดัชนีถ้อยคำสำคัญหลัก K บ่อยๆ หรือเป็นกลุ่มคำมีความคล้ายคลึงกันมาก

ในการวัดค่ากฎของความสัมพันธ์นั้นจะใช้วิธีการวัดค่าความเชื่อมั่น (confidence: c) คืออัตราส่วนระหว่างจำนวนครั้งที่พบว่าดัชนีถ้อยคำ ki อยู่กับดัชนีถ้อยคำสำคัญหลัก K จะเรียกว่า ค่าสนับสนุน (support) กับจำนวนดัชนีถ้อยคำ ki ทั้งหมดดังสมการที่1 (Mahgoub et al. 2008; Ding. 2002)

$$\text{confidence}(ki,K) = \frac{(\text{Support } (k_i,K)}{(\text{Support } (k_i)} \qquad \text{สมการที่ 1}$$

2) การวิเคราะห์การกระจายตัวเชิงพื้นที่ การวิเคราะห์การกระจายตัวเชิงพื้นที่จากการระบุตำแหน่งใช้เทคนิคการวิเคราะห์ความหนาแน่นของเคอร์เนล (kernel density) เพื่อสร้างข้อมูลกริดความหนาแน่นจากข้อมูลที่มีการระบุตำแหน่ง เพื่อใช้ในการหาพื้นที่ที่มีปริมาณความหนาแน่น (heatmap) หรือจุดฮอตสปอตจากการใช้งาน flickr โดยมีการกำหนดรัศมีเพื่อคำนวณความหนาแน่นให้เหมาะสมในแต่ละพื้นที่ ผลลัพธ์ที่ได้คือแผนที่แสดงความหนาแน่นต่ำสุดและสูงสุดและการกระจายตัวของข้อมูลการใช้งาน flickr แต่ละพื้นที่ตั้งแต่ปี 2010 – 2018 ดังแสดงในภาพที่ 4 ในส่วนของการวิเคราะห์เชิงพื้นที่จะเป็นการสกัดปริมาณข้อมูลในแต่ละช่วงเวลาและพื้นที่เพื่อนำมาแสดงถึงแนวโน้มการใช้งาน flickr ต่อไป



ภาพที่ 4: การวิเคราะห์ความหนาแน่นเชิงพื้นที่ของเคอร์เนล (kernel density)

ผลการวิจัย

ผลการวิจัยของโครงการวิจัย ประกอบด้วย 2 ส่วนตามวัตถุประสงค์ดังนี้

1. การดึงข้อมูลดัชนีถ้อยคำและจัดทำคลังฐานข้อมูล

การพัฒนาโปรแกรมเพื่อทำการดึงข้อมูลดัชนีถ้อยคำและจัดทำคลังฐานข้อมูล flickr ระหว่างปี 2553 - 2560 โดยมีผลลัพธ์ตัวอย่างในการดึงข้อมูลดังภาพ และจัดเก็บเป็นฐานข้อมูลดังภาพที่ 5 มีชุดข้อมูลที่ผู้ใช้งานอัปโหลดทั้งสิ้นจำนวน 6,018,380 ชุดข้อมูล ซึ่งในคลังฐานข้อมูลประกอบไปด้วย หมายเลข ชื่อผู้ใช้งาน วันและเวลาที่อัปโหลดข้อมูล ดัชนีถ้อยคำที่อธิบายข้อมูล และพิกัดแสดงตำแหน่งของข้อมูลนั้น

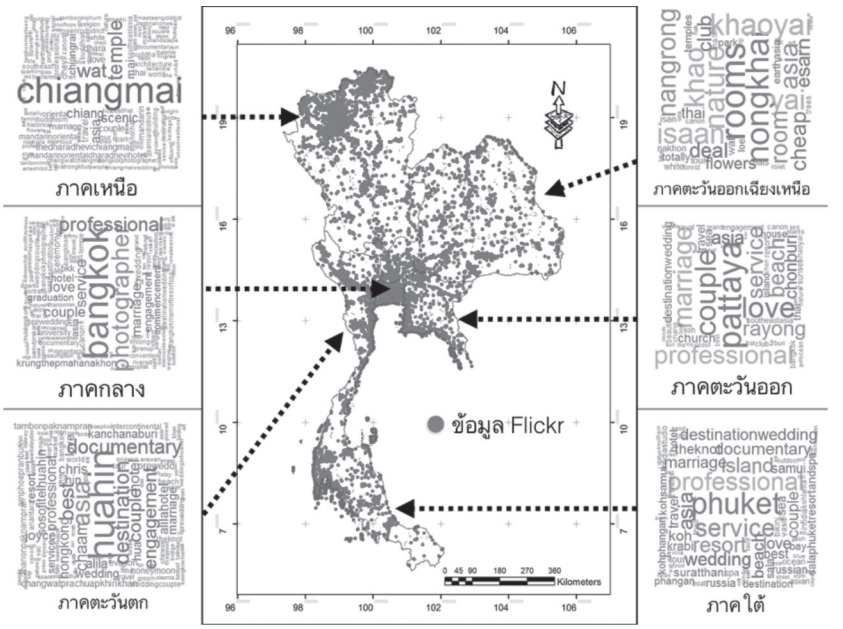
id	owner	datetaken	tags	latitude	longitude
6944245610	11784108@N06	2010-10-16 21:24:26	travel window thailand cabin asia vie...	11.416764	100.057296
6011287358	11784108@N06	2010-10-16 10:45:08	travel airplane thailand flying cabin a...	11.566143	100.437011
6056043111	42474633@N07	2010-10-10 13:35:43	travel thailand misc patthaya	12.682712	100.902729
5548667389	11784108@N06	2010-10-21 20:38:46	travel light night thailand hotel patta...	12.962404	100.890541
5548779477	11784108@N06	2010-10-21 20:41:58	travel thailand hotel mirage pattaya c...	12.964746	100.891227
6056590478	42474633@N07	2010-10-10 12:38:53	travel thailand misc patthaya	13.368450	100.987197
7941701348	68891011@N00	2010-10-19 14:42:29	travel cambodia angkor	13.441299	103.859746
5272795553	68891011@N00	2010-10-19 15:06:59	travel cambodia angkor	13.443062	103.856270
5200527807	58188719@N00	2010-10-24 17:27:50	travel boy face bike bicycle children p...	13.550382	103.944053
6105099969	11784108@N06	2010-10-16 08:45:31	travel thailand airport asia bangkok p...	13.684684	100.750293

ภาพที่ 5: ตัวอย่างคลังฐานข้อมูล flickr ระหว่างปี 2553 – 2560

2. การวิเคราะห์เหมืองดัชนีถ้อยคำ

2.1 การประมวลผลข้อความ

เทคนิคการประมวลผลข้อความจากดัชนีถ้อยคำ ประกอบไปด้วยขั้นตอน การเลือกข้อมูล การกรองข้อมูล และการจัดลำดับข้อมูล ให้อยู่ในรูปแบบคลังฐานข้อมูลเพื่อนำไปวิเคราะห์ พบว่า การกระจายของข้อมูล flickr ดังแสดงในภาพที่ 6 เมื่อพิจารณาภาพรวมระดับภูมิภาคของประเทศไทยแบ่งออกเป็นภาคกลาง ภาคใต้ ภาคตะวันตก ภาคตะวันออกและภาคตะวันออกเฉียงเหนือ นำข้อมูล flickr มาวิเคราะห์ความถี่ของดัชนีถ้อยคำที่เกิดขึ้นในรูปแบบถุงของข้อมูล (bag of words) และแสดงผลผ่านดัชนีถ้อยคำก้อนเมฆ (tagcloud) ในแต่ละภูมิภาค เช่น คำที่นิยมใช้ในการบรรยายในภาคเหนือคือ chaingmai, temple และ wat คำที่นิยมใช้ในการบรรยายในภาคใต้ คือ Phuket, island, koh และ samui เป็นต้น จากความนิยมในการระบุตำแหน่งของผู้ใช้งานพบว่าสถานที่ที่มีการระบุตำแหน่งส่วนใหญ่จะมีการกระจุกอยู่บริเวณสถานที่ เมืองหรือจังหวัดต่างๆ ที่มีความสำคัญ ซึ่งสามารถสะท้อนให้เห็นถึงเศรฐกิจ สังคม วัฒนธรรม และการท่องเที่ยวตามลักษณะภูมิประเทศ และจากผลการจัดอันดับจังหวัดที่มีความนิยมในการระบุตำแหน่งผ่าน flickr มากที่สุดจากการวิเคราะห์ความถี่ของดัชนีถ้อยคำ ได้แก่ กรุงเทพมหานคร เชียงใหม่ เชียงราย กระบี่ เพชรบุรี อยุธยา ภูเก็ต ประจวบคีรีขันธ์ สมุทรปราการและสุราษฎร์ธานี ตามลำดับ

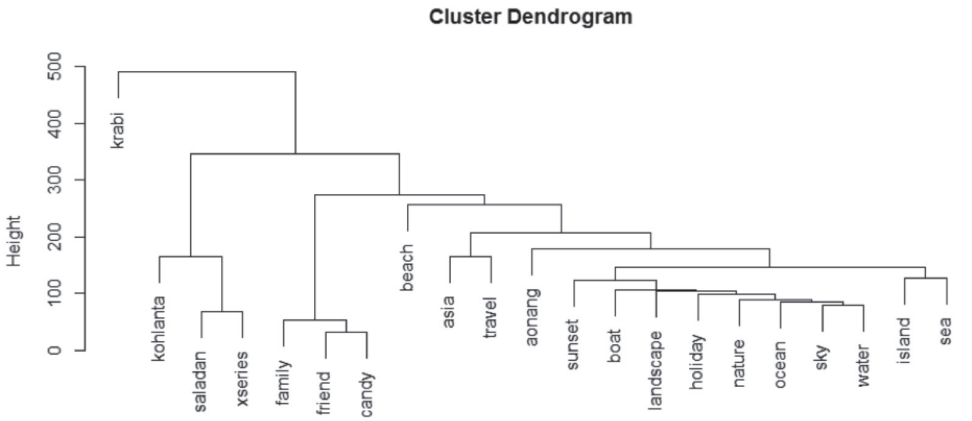


ภาพที่ 6: การกระจายของข้อมูลระบุตำแหน่งใน flickr และดัชนีถ้อยคำก้อนเมฆ (tagcloud) ในแต่ละภูมิภาคในประเทศไทย

2.2 การวิเคราะห์เหมืองดัชนีถ้อยคำ

2.2.1 การจัดกลุ่มแบบลำดับขั้น (Hierarchical Cluster Analysis)

การวิเคราะห์จัดกลุ่มข้อมูลโดยวิธีการจัดกลุ่มแบบลำดับขั้น (Hierarchical Cluster Analysis) ผลการวิจัยพบว่า ธรรมชาติของชุดข้อมูล flickr จะเป็นการตอบสนองการใช้งานในลักษณะการท่องเที่ยวตามสถานที่ต่างๆ การจัดกิจกรรม เป็นส่วนใหญ่ ภาพที่ 7 แสดงตัวอย่างแผนผังการจัดกลุ่ม (cluster dendrogram) ของดัชนีถ้อยคำจังหวัดกระบี่ จะเห็นได้ว่ากลุ่มดัชนีถ้อยคำที่มีความคล้ายคลึงกันนั้นเป็นข้อมูลประเภทการท่องเที่ยวเชิงธรรมชาติ อาทิ beach, sunset, ocean, sky, island และ sea ในส่วนของกลุ่มดัชนีถ้อยคำที่มีความคล้ายคลึงกันประเภทการการท่องเที่ยวเชิงวัฒนธรรม อาทิ temple, wat, art, และ buddhist ในส่วนของกลุ่มดัชนีถ้อยคำที่มีความคล้ายคลึงกันประเภทกิจกรรมและอื่น ๆ อาทิ graduation, marriage และ spa เป็นต้น จากการจัดกลุ่มข้อมูลโดยใช้เทคนิคลำดับขั้นในพื้นที่จังหวัดที่มีความนิยมใช้ flickr นั้น สามารถสรุปได้ว่าลักษณะข้อมูลที่มีการระบุตำแหน่งจะแบ่งออกเป็น ประเภทการท่องเที่ยวเชิงธรรมชาติ การท่องเที่ยวเชิงวัฒนธรรม และกิจกรรมและอื่น ๆ ซึ่งสอดคล้องกับลักษณะภูมิประเทศของจังหวัดกระบี่และสะท้อนถึงความนิยมในสถานที่ต่างๆ เป็นจำนวนมาก เป็นต้น



ภาพที่ 7: แผนผังการจัดการจัดกลุ่มข้อมูล

2.2.2 การหาความสัมพันธ์ระหว่างข้อมูล

1) กฎของการหาความสัมพันธ์ (Association Rule Mining: ARM)

ในงานวิจัยนี้ผู้วิจัยต้องการค้นหาความสัมพันธ์ระหว่างดัชนีถ้อยคำ (tags) จากข้อมูลการระบุตำแหน่งของผู้ใช้งาน flickr ซึ่งกำหนดให้เงื่อนไขกฎของความสัมพันธ์คือ กลุ่มดัชนีถ้อยคำสำคัญ (K) ผลลัพธ์ที่เกิดขึ้น คือ กลุ่มดัชนีถ้อยคำต่างๆ (k) ที่เกี่ยวข้องกับดัชนีถ้อยคำสำคัญ การประเมินค่าของกฎจะใช้ค่าความเชื่อมั่น (confidence) ของข้อมูลที่มีเงื่อนไขและผลลัพธ์สอดคล้องตามกฎของความสัมพันธ์ จากผลการประมวลผลข้อความในการจัดอันดับและวิเคราะห์ความดัชนีถ้อยคำที่พบว่า กลุ่มดัชนีถ้อยคำสำคัญ (K) ถูกแบ่งออกเป็น 3 ประเภทได้แก่ ประเภทกิจกรรมและอื่นๆ ประเภทการท่องเที่ยวเชิงวัฒนธรรมและประเภทการท่องเที่ยวเชิงธรรมชาติ ซึ่งสามารถเขียนตัวอย่างกฎความสัมพันธ์และแสดงค่าความเชื่อมั่นได้ ดังภาพที่ 8 ดังนี้

กลุ่มดัชนีถ้อยคำสำคัญ (K) ประเภทกิจกรรม กรุงเทพมหานคร เช่น กลุ่ม couple marriage และ service

กลุ่มดัชนีถ้อยคำสำคัญ (K) ประเภทการท่องเที่ยวเชิงวัฒนธรรม จังหวัดเชียงใหม่ เช่น กลุ่ม temple, wat และ scenenic

กลุ่มดัชนีถ้อยคำสำคัญ (K) ประเภทการท่องเที่ยวเชิงวัฒนธรรม จังหวัดกระบี่ เช่น กลุ่ม island, beach และ family

จากตัวอย่างความสัมพันธ์ข้างต้นนี้สามารถอธิบายได้ว่า กลุ่มดัชนีถ้อยคำต่างๆ (k) มีแนวโน้มที่จะพบหรือมีการบรรยายร่วมกับกลุ่มดัชนีถ้อยคำสำคัญ (K) ตามค่าความเชื่อมั่น (Confidence)

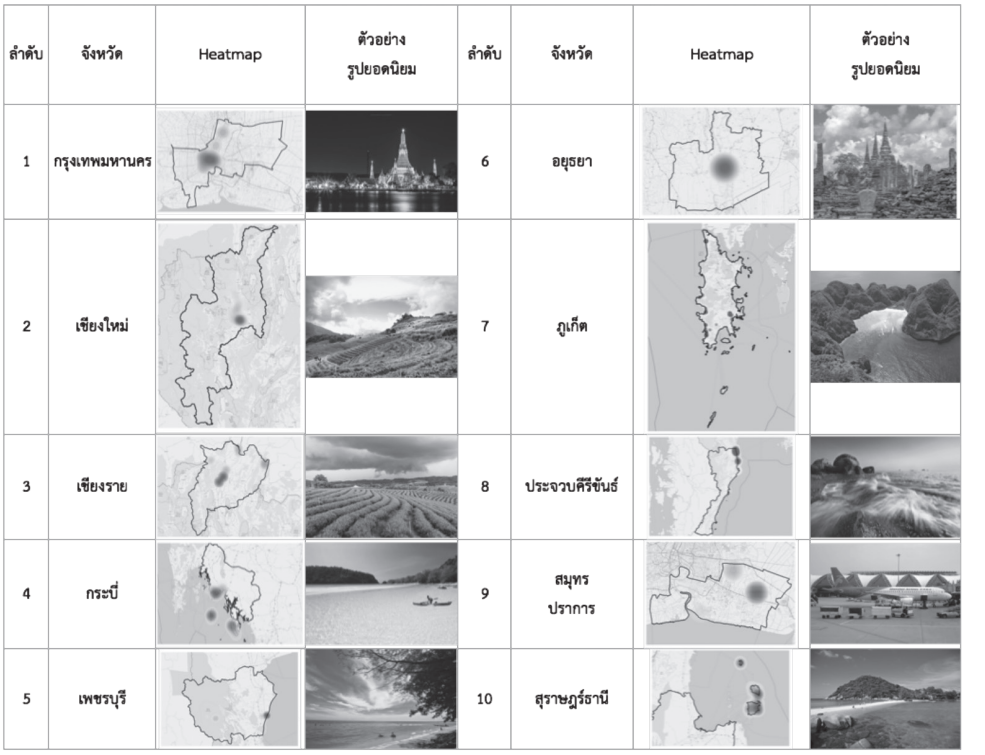
1. กรุงเทพมหานคร									
คำที่เกี่ยวข้องกับคำสำคัญ (k-keywords) และค่ากฎความสัมพันธ์กับคำสำคัญ (Associate's rule)									
bangkok		couple		love		marriage		service	
bangkokweddingphotor	0.56	love	0.82	marriage	0.99	love	0.99	professional	0.75
bangkokwedding	0.47	marriage	0.82	couple	0.82	couple	0.82	wedding	0.6
world	0.46	hotel	0.61	hotel	0.63	hotel	0.64	bkk	0.49
tour	0.43	ceremony	0.6	ceremony	0.61	ceremony	0.61	photographer	0.47
photographer	0.36	photographer	0.58	wedding	0.6	wedding	0.61	theknot	0.41
destination	0.32	wedding	0.56	professional	0.59	professional	0.59	love	0.4
bangkokphotographer	0.31	professional	0.51	photographer	0.48	destinationwedding	0.49	marriage	0.4
caturday	0.29	anantara	0.45	destinationwedding	0.48	photographer	0.48	weddingcouple	0.38
best	0.28	engagement	0.43	anantara	0.46	anantara	0.46	university	0.36
documentary	0.27	destinationwedding	0.43	service	0.4	resort	0.41	intercontinentalbangkok	0.34
2. เชียงใหม่									
คำที่เกี่ยวข้องกับคำสำคัญ (k-keywords) และค่ากฎความสัมพันธ์กับคำสำคัญ (Associate's rule)									
chiangmai		temple		wat		scenic		couple	
scenic	0.27	wat	0.86	temple	0.86	wat	0.81	dhara	0.99
temple	0.26	scenic	0.77	scenic	0.81	temple	0.77	dhevi	0.99
		maerimdistrict	0.38	maerimdistrict	0.4	maerimdistrict	0.49	thedharadhevichiangmai	0.99
		sansaidistrict	0.37	sansaidistrict	0.4	sansaidistrict	0.47	love	0.98
		tambonsriphum	0.29	tambonsriphum	0.3	tambonsriphum	0.37	mandarin	0.97
		maetaengdistrict	0.29	maetaengdistrict	0.3	maetaengdistrict	0.37	marriage	0.97
		mueangchiangmaidistrict	0.28	mueangchiangmaidistrict	0.29	mueangchiangmaidistrict	0.36	destinationwedding	0.96
		tambonmaefaekmai	0.27	tambonmaefaekmai	0.29	tambonmaefaekmai	0.34	oriental	0.94
		chiangmai	0.26	tambonmaetaeng	0.27	tambonmaetaeng	0.33	mandarinoriental	0.89
		tambonmaetaeng	0.26			tambonsanpong	0.3	mandarinorientaldharadhevihotel	0.89
3. กระบี่									
คำที่เกี่ยวข้องกับคำสำคัญ (k-keywords) และค่ากฎความสัมพันธ์กับคำสำคัญ (Associate's rule)									
asia		beach		family		island		kohlanta	
blue	0.4	ocean	0.36	candy	0.86	boat	0.49	saladan	0.76
bamboo	0.38	resort	0.32	friend	0.79	relax	0.47	xseries	0.66
bay	0.38	island	0.31	travel	0.53	blue	0.45	relaxbay	0.35
water	0.38	sky	0.31	bamboo	0.45	sea	0.45	krabi	0.31
relax	0.38	water	0.29	landscape	0.45	bay	0.45		
laut	0.38	holiday	0.29	laut	0.45	tour	0.45		
elephant	0.37	relax	0.29	elephant	0.43	ocean	0.44		
wanderlust	0.37	blue	0.28	sigma	0.42	resort	0.44		
island	0.36	sea	0.27	tour	0.41	bamboo	0.43		
travel	0.36	nature	0.27	wanderlust	0.41	nature	0.42		

ภาพที่ 8: กฎของการหาความสัมพันธ์ระหว่างดัชนีถ้อยคำจากข้อมูล flickr

2) การวิเคราะห์การกระจายตัวเชิงพื้นที่และเชิงเวลา

การใช้เทคนิคการวิเคราะห์ความหนาแน่นของเคอร์เนล (kernel density) ในการหาความหนาแน่นของผู้ใช้งาน flickr โดยกำหนดรัศมีที่แตกต่างกันตามขนาดพื้นที่ในแต่ละจังหวัด จากแผนที่ความหนาแน่น (heatmep) ที่วิเคราะห์ได้ พบว่าสีแดงเป็นบริเวณที่มีการระบุตำแหน่งผ่าน flickr หนาแน่นมากและหนาแน่นรองลงมาจะแสดงด้วยระดับสีที่อ่อนลงตามลำดับโดยการแบ่งชั้นข้อมูลความหนาแน่นด้วยวิธีการแบ่งกลุ่มของข้อมูลแบบธรรมชาติ (natural brake) ตามลักษณะของข้อมูล ซึ่งเป็นตัวกำหนดค่าที่คล้ายคลึงกันของกลุ่มได้ดีที่สุดและความแตกต่างระหว่างแต่ละชั้นข้อมูล ภาพที่ 9 แสดงตัวอย่างแผนที่ความหนาแน่นของข้อมูลระบุตำแหน่งใน flickr และตัวอย่างภาพที่ได้รับความนิยมสูงสุด 10 จังหวัดในประเทศไทย กล่าวได้ว่า ข้อมูลผู้ใช้งานมักจะกระจุกตัวบริเวณ

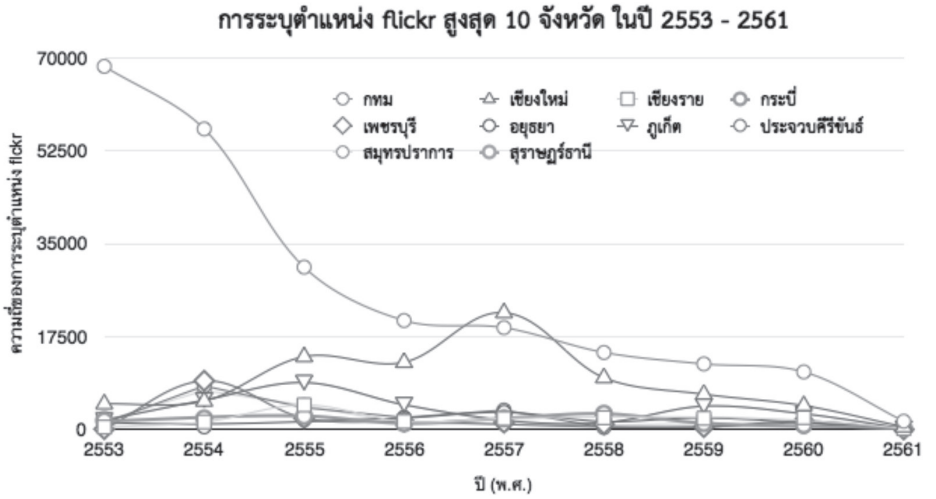
ที่เป็นสถานที่ที่มีเอกลักษณ์ในการทำกิจกรรมและสถานที่ท่องเที่ยวต่างๆ ที่อาจมีการเข้าถึงทางด้านคมนาคมขนส่งหรือมีโครงสร้างพื้นฐานที่ดี จึงได้รับความนิยมในการระบุตัวตนจากข้อมูล flickr ที่มีการแสดงความคิดเห็นในทางบวกเกี่ยวกับสถานที่นั้นๆ เป็นส่วนใหญ่ เช่นคำว่า best, beautiful, relax และ love



ภาพที่ 9: แผนที่ความหนาแน่น (Heatmap) ของข้อมูลระบุตำแหน่งใน flickr และตัวอย่างภาพที่ได้รับความนิยมสูงสุด 10 จังหวัดในประเทศไทย

ในส่วนของการวิเคราะห์ข้อมูลการใช้ flickr เชิงเวลา ภาพที่ 10 แสดงการเปรียบเทียบข้อมูล flickr ตั้งแต่ปี 2553 ถึงปัจจุบันพบว่า กรุงเทพมหานครมีการใช้งาน flickr จำนวนสูงสุดโดยเฉพาะในปี 2551 เนื่องมาจากเป็นพื้นที่ศูนย์กลางทางด้านเทคโนโลยีและเศรษฐกิจ อย่างไรก็ตามกรุงเทพมหานครมีการใช้งานที่ลดลงอย่างต่อเนื่อง จังหวัดเชียงใหม่มีการใช้งานเป็นอันดับที่ 2 และเพิ่มขึ้นสูงสุดในปี 2557 อาจมีส่วนมาจากการได้รับความนิยมของนักท่องเที่ยวชาวต่างชาติ และมีแนวโน้มการใช้งานที่ลดลงเล็กน้อยและคงที่จนถึงปัจจุบัน ในขณะที่จังหวัดอื่น ๆ ได้แก่ เชียงราย กระบี่ เพชรบุรี อยุธยา ภูเก็ต ประจวบคีรีขันธ์ สมุทรปราการ และสุราษฎร์ธานี มีแนวโน้มการใช้งานไม่เปลี่ยนแปลงมากนัก อย่างไรก็ตาม เมื่อพิจารณาจากปริมาณการระบุตำแหน่ง flickr สูงสุด 10 จังหวัด ในปี 2553 – 2561 (เดือนมกราคมจนถึงเดือนเมษายน) (ภาพที่ 10) กล่าวได้ว่า การใช้ flickr ในการแบ่งปันข้อมูลในภาพรวมอาจมีแนวโน้มและการเปลี่ยนแปลงไม่มากนัก แต่เมื่อพิจารณาการใช้งานในช่วงเวลาตามฤดูกาลนั้น จะมีการสอดคล้องกับการกระจายตัวเชิงพื้นที่จากการใช้งาน flickr

เช่น จังหวัดเชียงใหม่มีปริมาณการใช้งาน flickr ในช่วงเดือนพฤศจิกายนถึงเดือนมกราคมเป็นจำนวนมาก ซึ่งสะท้อนให้เห็นความนิยมในการท่องเที่ยวตามลักษณะภูมิประเทศและภูมิอากาศ เป็นต้น ด้วยเทคโนโลยีที่ก้าวหน้าและทันสมัยทำให้มีอุปกรณ์ที่สามารถใช้ได้ทั้งถ่ายภาพและสื่อสารไร้สายอย่างรวดเร็ว ประกอบกับมีการพัฒนาแพลตฟอร์มจำนวนมากในการได้มาซึ่งข้อมูลข่าวสารและกิจกรรมต่าง ๆ flickr เป็นแพลตฟอร์มทางเลือกหนึ่งในการแบ่งปันรูปภาพสาธารณะจากกลุ่มผู้รักการถ่ายภาพและมีความชื่นชอบในการใช้งาน flickr และยังได้รับความนิยมในการใช้จนถึงปัจจุบัน



ภาพที่ 10: การระบุตำแหน่ง flckr สูงสุด 10 จังหวัด ในปี 2553 - 2561

สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์เหมืองดัชนีข้อมูลเชิงตำแหน่งผ่านสื่อสังคมออนไลน์ กรณีศึกษาแพลตฟอร์ม flickr โดยแบ่งการวิเคราะห์ออกเป็น 2 ขั้นตอนหลัก ได้แก่ การพัฒนาคลังฐานข้อมูลและการวิเคราะห์เหมืองดัชนีถ้อยคำเพื่อหาความสัมพันธ์ของข้อมูลทั้งในเชิงพื้นที่และเชิงเวลา ดัชนีถ้อยคำและพิกัดที่มีการระบุตำแหน่งถูกนำมาเป็นข้อมูลสำคัญที่ใช้ในการวิเคราะห์เหมืองดัชนีถ้อยคำ เทคนิคการวิเคราะห์เหมืองดัชนีถ้อยคำประกอบไปด้วย 1) การประมวลผลข้อความจากคลังฐานข้อความ ได้แก่ การเลือกคุณลักษณะของข้อมูล การกรองข้อมูลและการจัดลำดับสำคัญของข้อมูลเพื่อนำไปสู่ 2) การวิเคราะห์เหมืองดัชนีถ้อยคำ โดยการจัดกลุ่มข้อมูลแบบลำดับขั้น (Hierarchical Cluster analysis) กฎของการหาความสัมพันธ์ (Association Rule Mining) และการวิเคราะห์การกระจายตัวเชิงพื้นที่และเชิงเวลา

ผลการวิจัยพบว่า ลักษณะข้อความที่มีการบรรยายร่วมกับการระบุตัวตนในแต่ละสถานที่ ซึ่งเมื่อพิจารณาร่วมกับผลการจัดกลุ่มข้อมูล ค่าความเชื่อมั่นจากกฎของความสัมพันธ์จากดัชนีถ้อยคำสำคัญ และแผนที่ความหนาแน่นของข้อมูลระบุตำแหน่ง (heatmap) เชิงพื้นที่ พบว่า พฤติกรรมหรือลักษณะการใช้งาน flickr สามารถจัดกลุ่มได้ 3 ประเภท ได้แก่ ประเภทกิจกรรมและอื่นๆ ประเภทการท่องเที่ยวเชิงวัฒนธรรม และประเภทการท่องเที่ยวเชิงธรรมชาติ ซึ่งมีความสอดคล้องกันอย่างเห็นได้ชัด

ทั้งนี้ยังสะท้อนให้เห็นถึงข้อความที่สื่อถึงกิจกรรมหรือลักษณะภูมิประเทศของสถานที่ที่มีการระบุตำแหน่งนั้นๆ อีกด้วย ข้อเสนอแนะจากผลการวิจัยครั้งนี้ พบว่าการวิเคราะห์เหมืองข้อความจากข้อมูลสื่อสังคมออนไลน์ สามารถนำมาวิเคราะห์ร่วมกับข้อมูลตัวอักษร รูปภาพ ที่มีการแบ่งปันผ่านสื่อสังคมออนไลน์ ที่มีการระบุตำแหน่งอื่น ๆ ได้แก่ twitter instagram, foursquare, panoramio หรือ openstreet view และการใช้เทคนิควิธีการวิเคราะห์อื่น ๆ เช่นการรู้จำและจำแนกประเภทของข้อมูลขั้นสูงจากปัญญาประดิษฐ์ (Artificial Intelligence) ทำให้สามารถสะท้อนพฤติกรรมของการใช้งานและขั้วความคิดเห็น ซึ่งจะช่วยส่งเสริมในการวางแผน พัฒนา ทั้งในด้านธุรกิจ การท่องเที่ยว การคมนาคม และโครงสร้างพื้นฐานต่างๆ รวมไปถึงการประยุกต์ในด้านการรายงานเหตุการณ์ การเตือนภัยและการเฝ้าระวัง เช่น สภาพการจราจร การเกิดอุบัติเหตุ หรือภัยพิบัติต่างๆ ในเชิงพื้นที่ให้ดียิ่งขึ้น การใช้เทคโนโลยีสื่อสังคมสำหรับภูมิสารสนเทศศาสตร์ เป็นเทคโนโลยีที่น่าสนใจอีกทางเลือกหนึ่ง เพื่อให้ทราบถึงลักษณะการรับรู้ทางสังคม จากการได้มาซึ่งข้อมูลสาธารณะซึ่งจะเป็นประโยชน์สำหรับการใช้งานและวางแผนการจัดการเชิงพื้นที่ในด้านโครงสร้างพื้นฐาน การคมนาคมขนส่งในอนาคตได้อย่างเหมาะสมและมีประสิทธิภาพต่อไป

เอกสารอ้างอิง

กานดา แผ่วัฒนากุล. และ ปราโมทย์ ลือนาม. (2556) การวิเคราะห์เหมืองความคิดเห็นบนเครือข่ายสังคมออนไลน์. **วารสารการจัดการสมัยใหม่.** 11(2): 11.

คณะกรรมการภูมิสารสนเทศแห่งชาติ. (2559). **แผนแม่บทภูมิสารสนเทศแห่งชาติ พ.ศ.2560-2564.** กรุงเทพฯ: คณะกรรมการภูมิสารสนเทศแห่งชาติ กระทรวงวิทยาศาสตร์และเทคโนโลยี (วท.).

เอกสิทธิ์ พัชรวงศ์ศักดา. (2557). **การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไมน์นิง เบื้องต้น.** กรุงเทพฯ. เอเชีย ดิจิตอลการพิมพ์

Ding, Q., Perrizo, W. (2002). **Association Rule Mining on Remotely Sensed Images Using P-trees. In: Chen MS., Yu P.S., Liu B. (eds) Advances in Knowledge Discovery and Data Mining. PAKDD 2002.** Lecture Notes in Computer Science, vol 2336, Berlin: Springer.

Mahgoub, H., Rösner, D.F., Ismail, N.F., & Torkey, F.A. (2008). **A Text Mining Technique Using Association Rules Extraction.**

Sergieh H.M.., Gabriele, G., Mario, D., Harald, K., Elöd, E. & Jean-Marie, P. (2012). **Geo-based automatic image annotation.** In Proceedings of the 2nd ACM International Conference on Multimedia Retrieval (ICMR ‘12). ACM, New York, NY, USA, Article 46, 8 pages. DOI: https://doi.org/10.1145/2324796.2324850.

Huang, H., Gartner, G. (2018). Current Trends and Challenges in Location-Based Services. **ISPRS Int. J. Geo-Inf.** 7: 199.

Huang, H., Gartner, G. (2016). Using mobile crowdsourcing and geotagged social media data to study people’s affective responses to environments. In: Capineri, Cristina; Haklay, Muki; Huang, Haosheng; Antoniou, Vyron; Kettunen, Juhani; Ostermann, Frank O; Purves, Ross S. **European Handbook of Crowdsourced Geographic Information.** London: Ubiquity Press, 385-399.

Murtagh, F. (2009) Symmetry in data mining and analysis: a unifying view based on hierarchy. **Proceedings of Steklov Institute of Mathematics 2009,** 265: 177–198.

Jomsri, P. (2012). Analyzing the Relation of Community Group for Research Paper Bookmarking by Using Association Rule. **World Academy of Science, Engineering and Technology.** 69: 184.

Richmond and Riva. (2010). Three Best Ways to Use Location-Based Social Media. **The Wall Street Journal.** New York: Dow Jones & Company, Inc.