

วิธีการสกัดข่าวด้านความปลอดภัย

A Method for Safety-Related News Extraction

สุรจิต ภูมิคง ธรา อังสกุล และจิติมนต์ อังสกุล*

Surajit Pumikong Thara Angskun and Jitimon Angskun *

สาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี

Abstract

Safety is an important factor to tourist's decision for travel planning because there is currently so much violence in the world. The violence may be caused by natural disasters, political, or social unrest. Thus, an online travel planner under a safety constraint was introduced. The travel planner applied safety-related news or events that occurred during a specified time period for travel planning. The administrator must search, analyze and update news by themselves. However, if news is obsolete, the travel plan generation may cause errors. This research aims to develop a safety-related news extraction method for supporting the online travel planner under a safety constraint. The method facilitates the travel planner to analyze and update news automatically. It retrieves safety-related news from news sites and classifies them into five categories: Health, Traffic, Natural disaster, Country's unrest and Crime. Then the method extracts safety-related keywords from the contents of news in order to apply for analyzing the level of violence of events and predicting the time to finish the events. The experimental results illustrate that the proposed method achieves with 82.84% of accuracy in extracting the safety-related keywords and 0.35 of mean absolute error (MAE) in predicting the level of violence.

Keywords: News Extraction; Safety-Related News; Level of Violence

บทคัดย่อ

ความปลอดภัยนับเป็นปัจจัยที่สำคัญอย่างยิ่งซึ่งส่งผลต่อการตัดสินใจของนักท่องเที่ยวในการวางแผนการเดินทาง เนื่องจากในปัจจุบันมีเหตุการณ์ความรุนแรงเกิดขึ้นทั่วโลกบ่อยครั้ง ทั้งจากภัยธรรมชาติ และจากความไม่สงบทางการเมือง หรือทางสังคม ดังนั้นจึงมีผู้คิดค้นระบบวางแผนการท่องเที่ยวออนไลน์ภายใต้เงื่อนไขบังคับด้านความปลอดภัยขึ้นมา แต่อย่างไรก็ตาม ระบบดังกล่าวเป็นการนำเอาข่าวสารหรือเหตุการณ์ที่เกี่ยวข้องกับความปลอดภัยที่เกิดขึ้นในช่วงเวลานั้น ๆ มาเป็นตัวแปรในการวิเคราะห์แผนการเดินทางท่องเที่ยว ซึ่งผู้ดูแลระบบต้องเป็นผู้ค้นหาข่าวสาร วิเคราะห์ข้อมูล และปรับปรุงข้อมูลด้วยตนเอง ซึ่งหากไม่มีการปรับปรุงข้อมูลดังกล่าว จะส่งผลให้การวางแผนการเดินทางท่องเที่ยวผิดพลาดได้ ในงานวิจัยนี้จึงมุ่งพัฒนาวิธีการสกัดข่าวด้านความปลอดภัยแบบอัตโนมัติ เพื่อสนับสนุนระบบวางแผนการท่องเที่ยวออนไลน์ภายใต้เงื่อนไขบังคับด้านความปลอดภัย ให้สามารถวิเคราะห์และปรับปรุงข่าวสารด้านความปลอดภัยให้ทันสมัยอยู่เสมอ วิธีการนี้ทำโดยการดึงข่าวที่เกี่ยวข้องกับความปลอดภัยจากเว็บไซต์รวบรวมข่าว และจัด

* ผู้เขียนที่ให้การติดต่อ โทร. +66 4422 4309; Fax: +66 4422 4205
E-mail address: jitimon@sut.ac.th

ประเภทของข่าวเป็น 5 ด้าน ได้แก่ ด้านสุขภาพ ด้านการจราจร ด้านภัยธรรมชาติ ด้านความไม่สงบของบ้านเมือง และด้านอาชญากรรม หลังจากนั้นคำศัพท์ที่เกี่ยวข้องกับความปลอดภัยจะถูกสกัดออกจากเนื้อข่าว เพื่อนำไปใช้ในการวิเคราะห์ระดับความรุนแรงของเหตุการณ์ และคาดการณ์ระยะเวลาสิ้นสุดของเหตุการณ์ ซึ่งจากผลการทดลองพบว่าวิธีการที่นำเสนอนี้มีค่าความถูกต้องในการสกัดคำศัพท์ที่เกี่ยวข้องกับความปลอดภัย เท่ากับ 82.84% และมีค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ในการคาดการณ์ระดับความรุนแรงของเหตุการณ์ เท่ากับ 0.35

บทนำ

ในปีที่ผ่านมา ถึงแม้ว่าประเทศไทยจะได้รับผลกระทบจากเหตุการณ์รุนแรงภายในประเทศ ทั้งจากภัยธรรมชาติ และจากมนุษย์ด้วยกันเอง แต่กระทรวงการท่องเที่ยวและกีฬา และการท่องเที่ยวแห่งประเทศไทยก็มีการดำเนินมาตรการเพื่อสร้างความมั่นใจและกระตุ้นการท่องเที่ยวของประเทศไทย และจากการสรุปสถานการณ์การท่องเที่ยวของประเทศไทยในปี พ.ศ. 2553 พบว่ามีจำนวนนักท่องเที่ยวประมาณ 15.84 ล้านคน จากอัตราการขยายตัวในระดับสูงของนักท่องเที่ยว เมื่อเทียบกับปี พ.ศ. 2552 (กระทรวงการท่องเที่ยวและกีฬา, 2553) จากข้อมูลดังกล่าว แสดงให้เห็นถึงความมั่นใจในความปลอดภัยของนักท่องเที่ยวที่เพิ่มมากขึ้น ประกอบกับการส่งเสริมการท่องเที่ยวของรัฐบาล โดยเฉพาะปัจจัยด้านความปลอดภัยที่เป็นปัจจัยสำคัญที่ทำให้นักท่องเที่ยวตัดสินใจเดินทางมาท่องเที่ยวในประเทศไทย ซึ่งสอดคล้องกับการศึกษาของ ชาง คิม และตวง (Zhang, Qu and Tang, 2003) ที่พบว่าความปลอดภัยเป็นมิติที่สำคัญที่สุดสำหรับผู้เลือกสถานที่ท่องเที่ยวพักผ่อน ซึ่งการท่องเที่ยวแห่งประเทศไทยก็ตระหนักถึงความสำคัญของความปลอดภัยของนักท่องเที่ยว

ความไม่ปลอดภัยในการท่องเที่ยวเกิดขึ้นจากหลายสาเหตุ เช่น การเกิดน้ำท่วมฉับพลัน การเกิดอุบัติเหตุรถชน และการเกิดจลาจล เป็นต้น ซึ่งเหตุผลเหล่านี้ย่อมส่งผลต่อการเดินทางท่องเที่ยวเป็นอย่างมาก และจากการสำรวจระบบที่ให้บริการค้นหาข้อมูลแหล่งท่องเที่ยว และบริการวางแผนการท่องเที่ยวแบบต่าง ๆ ที่มีอยู่ในปัจจุบัน พบว่า ได้มีผู้พัฒนาระบบวางแผนท่องเที่ยวออนไลน์ภายใต้เงื่อนไขบังคับด้านความปลอดภัยขึ้นมาแล้ว (สุวรรณ บุษเล และคณะ, 2554) โดยระบบดังกล่าวเป็นการวางแผนการท่องเที่ยวโดยคำนึงถึงปัจจัยด้านความปลอดภัยเป็นหลัก และแบ่งปัจจัยที่เกี่ยวข้องกับความปลอดภัยในการท่องเที่ยวออกเป็น 5 ด้าน คือ ด้านสุขภาพ ด้านการจราจร ด้านภัยธรรมชาติ ด้านความไม่สงบของบ้านเมือง และด้านอาชญากรรม โดยระบบนี้เป็นการนำเอาข่าวสารที่เกี่ยวข้องกับความปลอดภัยมาเป็นตัวแปรในการวิเคราะห์การวางแผนการเดินทางท่องเที่ยว ซึ่งข่าวสารหรือเหตุการณ์ด้านความปลอดภัยที่เกิดขึ้นในแต่ละช่วงเวลานั้น ผู้ดูแลระบบจะเป็นผู้ค้นหาข่าวสาร วิเคราะห์ข้อมูล และปรับปรุงข้อมูลด้วยตนเอง ซึ่งหากไม่มีการปรับปรุงข้อมูลดังกล่าว ก็จะส่งผลให้การวางแผนการเดินทางท่องเที่ยวผิดพลาดได้ ดังนั้นจึงนำมาสู่แนวคิดในการพัฒนาวิธีการสกัดข่าวแบบอัตโนมัติ เพื่อสนับสนุนระบบวางแผนการท่องเที่ยวออนไลน์ภายใต้เงื่อนไขบังคับด้านความปลอดภัยขึ้นมา โดยการนำเอาข่าวที่เกี่ยวข้องกับความปลอดภัยมาเป็นปัจจัยหลักในการพัฒนาเพื่อให้ระบบดังกล่าวสามารถวิเคราะห์และปรับปรุงข่าวสารด้านความปลอดภัยให้ทันสมัยอยู่เสมอ

ทำให้ระบบวางแผนการท่องเที่ยวมีประสิทธิภาพมากยิ่งขึ้น

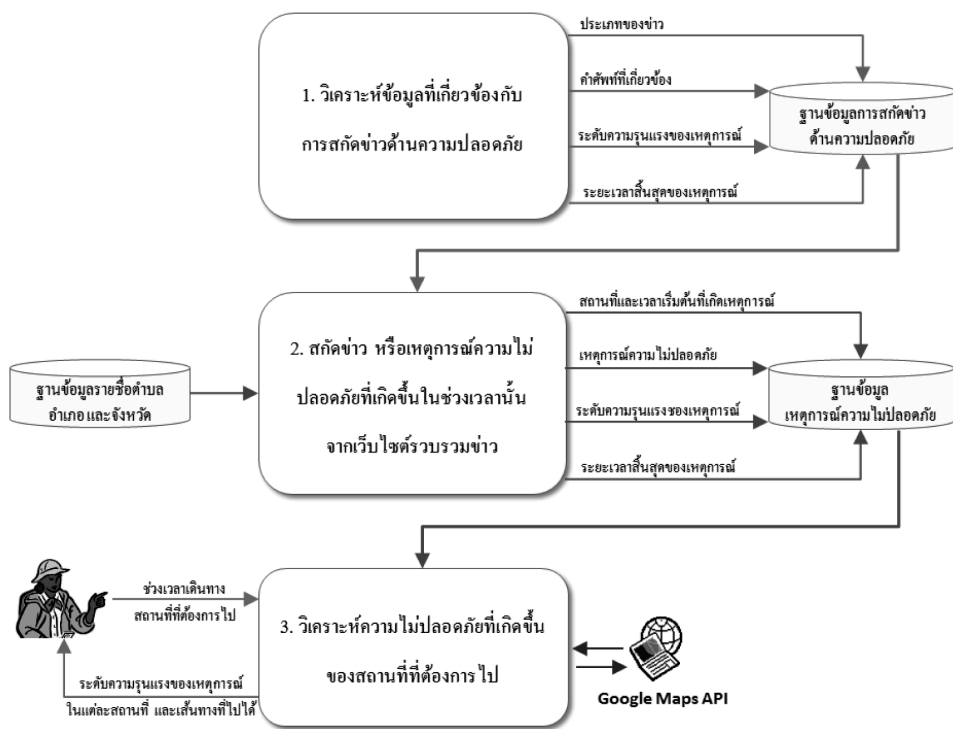
ในการทบทวนวรรณกรรมที่เกี่ยวข้องกับวิธีการสกัดข่าว ได้แบ่งการศึกษาออกเป็น 2 ด้าน ได้แก่ การศึกษางานวิจัยที่เกี่ยวข้องกับการตัดคำ เพื่อทำความเข้าใจหลักการทำงานของการตัดคำหรือการสกัดคำออกจากบทความหรือวารสารที่มีเนื้อหาที่ต้องการ และการศึกษางานวิจัยที่เกี่ยวข้องกับการสกัดสารสนเทศ เพื่อใช้เป็นแนวทางในการสกัดสารสนเทศจากข่าวออนไลน์ที่ต้องการ

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการตัดคำ พบว่า งานวิจัยส่วนใหญ่เป็นการประยุกต์หลักการต่าง ๆ เพื่อปรับใช้กับหลักการตัดคำโดยใช้พจนานุกรม ทั้งการปรับปรุงพจนานุกรมให้กะทัดรัดขึ้น เพิ่มคำที่ไม่มีในพจนานุกรมให้ครอบคลุมขึ้นและลดความกำกวมของคำในพจนานุกรม โดยมุ่งเน้นให้การใช้พจนานุกรมเกิดประสิทธิภาพมากที่สุด ซึ่งต่างจากหลักการของการใช้กฎ (Rule Base) (นัฐวุฒิ ไชยเจริญ, 2544) และหลักการใช้เอ็นแกรม (n-gram) (Park, Tae, and Park, 2006) เนื่องจากหลักการใช้กฎและเอ็นแกรมมีหลักการทำงานที่ซับซ้อนและยากที่จะทำให้ครอบคลุมคำทั้งหมดได้จึงทำให้การตัดคำโดยใช้พจนานุกรมมีประสิทธิภาพการทำงานที่ดีกว่า และเป็นที่ยอมรับกันโดยทั่วไป

สำหรับการศึกษางานวิจัยที่เกี่ยวข้องกับการสกัดสารสนเทศ พบว่า งานวิจัยส่วนใหญ่ที่พบเป็นการสกัดข่าวโดยทั่วไปไม่เฉพาะเจาะจงว่าเป็นข่าวด้านใดด้านหนึ่ง อาทิ การสกัดเฉพาะข่าวกีฬา การสกัดเฉพาะข่าวบันเทิง ทำให้ในการสกัดสารสนเทศออกมาจำเป็นต้องมุ่งเน้นไปที่การกำจัดสิ่งรบกวนต่าง ๆ ที่อยู่บนหน้าเว็บ ส่งผลให้การสกัดข่าวไม่มีประสิทธิภาพเท่าที่ควร แต่ก็มีบางงานวิจัยที่เป็นการสกัดข่าวแบบเฉพาะเจาะจง คือ งานวิจัยของฮัน และคณะ (Han, Noro and Tokuda, 2009) ซึ่งเป็นการสกัดข่าวเพื่อต้องการจำแนกข่าวออกตามแหล่งที่มาของข่าว และงานวิจัยของ หลาน (Lan, 2010) ซึ่งมีวัตถุประสงค์เพื่อเปรียบเทียบเว็บไซต์ 2 เว็บไซต์ที่มีหัวข้อข่าวคล้ายกันและมีการเผยแพร่ข้อมูลในเวลาใกล้เคียงกัน แต่อย่างไรก็ตาม กระบวนการที่ได้จากงานวิจัยทั้งสองงานนี้ไม่สามารถนำไปประยุกต์กับการสกัดข่าวที่เกี่ยวข้องกับความปลอดภัยที่งานวิจัยนี้นำเสนอได้ เนื่องจากงานวิจัยทั้งสองนั้นมีวัตถุประสงค์ของการสกัดสารสนเทศที่ต่างออกไปกัน ทำให้การนำเข้าข้อมูลกระบวนการ และผลลัพธ์ที่ได้มีความแตกต่างกันเชิง จึงไม่สามารถนำมาใช้งาน หรือเปรียบเทียบกันได้ ดังนั้นงานวิจัยนี้จึงมุ่งพัฒนาวิธีการสกัดข่าวด้านความปลอดภัยแบบอัตโนมัติ โดยการดึงข่าวที่เกี่ยวข้องกับความปลอดภัยจากเว็บไซต์รวบรวมข่าว นำมาจัดประเภท และวิเคราะห์คำศัพท์ที่เกี่ยวข้องกับความไม่ปลอดภัยนั้น เพื่อนำมาใช้เป็นฐานข้อมูลในการสกัดคำศัพท์ที่เกี่ยวข้องกับความไม่ปลอดภัยออกจากเนื้อหาข่าวแบบอัตโนมัติ โดยคำศัพท์ที่สกัดได้จะถูกนำไปใช้ในการวิเคราะห์ระดับความรุนแรงของเหตุการณ์ และคาดการณ์ระยะเวลาสิ้นสุดของเหตุการณ์ในแต่ละสถานที่ต่อไป โดยผลลัพธ์ที่ได้จากงานวิจัยนี้จะถูกนำไปใช้เพื่อสนับสนุนระบบวางแผนการท่องเที่ยวออนไลน์ภายใต้เงื่อนไขบังคับด้านความปลอดภัย ให้สามารถวิเคราะห์และปรับปรุงข่าวสารด้านความปลอดภัยให้ทันสมัยอยู่เสมอ

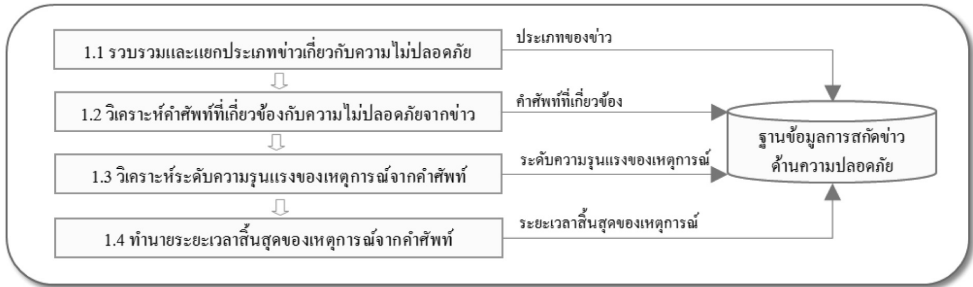
วิธีการดำเนินงานวิจัย

ในการออกแบบและพัฒนาวิธีการสกัดข่าวด้านความปลอดภัย มุ่งเน้นการออกแบบและพัฒนาขั้นตอนวิธีการสกัดข่าวภาษาไทยแบบอัตโนมัติ เพื่อสนับสนุนระบบการวางแผนการท่องเที่ยวออนไลน์ภายใต้เงื่อนไขบังคับด้านความปลอดภัย จากการศึกษางานวิจัยที่เกี่ยวข้อง (มหาวิทยาลัยสุโขทัยธรรมาราช, 2544; สุวรรณ บุเหลา และคณะ, 2554) พบว่า ปัจจัยที่เกี่ยวข้องกับความปลอดภัยสำหรับการท่องเที่ยว สามารถแบ่งได้เป็น 3 ประเภท ได้แก่ ความไม่ปลอดภัยในอดีต ความไม่ปลอดภัยในปัจจุบัน และการสนับสนุนด้านความปลอดภัย โดยงานวิจัยนี้จะมุ่งศึกษาเฉพาะความไม่ปลอดภัยในปัจจุบันเป็นสำคัญ ซึ่งความไม่ปลอดภัยในปัจจุบัน หมายถึง การเกิดเหตุการณ์ต่าง ๆ ที่อาจก่อให้เกิดความไม่ปลอดภัยในการท่องเที่ยว ณ ช่วงเวลาที่นักท่องเที่ยววางแผนเดินทางไปท่องเที่ยว ซึ่งหากเกิดเหตุการณ์รุนแรงก็ไม่ควรวางแผนการเดินทางให้ผ่านสถานานั้น เช่น การเกิดอุทกภัย การเกิดแผ่นดินไหว และการชุมนุมทางการเมือง เป็นต้น ซึ่งปัจจัยดังกล่าวถูกนำมาใช้ในการประเมินความปลอดภัยของแผนการท่องเที่ยว โดยปัจจัยด้านความไม่ปลอดภัยในปัจจุบันนั้น ได้มาจากข่าวเหตุการณ์ที่เกิดขึ้น ณ ช่วงเวลานั้น ดังนั้นในงานวิจัยนี้จึงมุ่งเน้นศึกษางานวิจัยที่เกี่ยวข้องรวบรวมข้อมูล และวิเคราะห์ปัจจัยที่เกี่ยวข้องกับข่าวด้านความปลอดภัย และออกแบบวิธีการสกัดข่าวแบบอัตโนมัติ โดยมีขั้นตอนการดำเนินงานดังภาพที่ 1



ภาพที่ 1 วิธีการสกัดข่าวด้านความปลอดภัย

ขั้นตอนที่ 1 วิเคราะห์ข้อมูลที่เกี่ยวข้องกับการสกัดข่าวด้านความปลอดภัย โดยการวิเคราะห์ข้อมูลที่เกี่ยวข้องกับการสกัดข่าวด้านความปลอดภัยนั้นมีการดำเนินการดังภาพที่ 2 ซึ่งมีภาพที่รายละเอียดดังต่อไปนี้



ภาพที่ 2 การวิเคราะห์ข้อมูลที่เกี่ยวข้องกับการสกัดข่าวด้านความปลอดภัย

1.1 รวบรวมและแยกประเภทข่าวเกี่ยวกับความไม่ปลอดภัย โดยในขั้นตอนนี้ได้เก็บรวบรวมข่าวที่เกี่ยวกับความไม่ปลอดภัยอาทิ ข่าวอาชญากรรม และข่าวประจำวัน เพื่อนำไปใช้เป็นฐานข้อมูลในการสกัดข่าว โดยได้เก็บรวบรวมข่าว ณ ช่วงเวลาที่ทำการวิจัย เป็นระยะ ๆ ตลอดระยะเวลาทั้งหมด 3 เดือน รวมทั้งสิ้น 150 ข่าว จากเว็บรวบรวมข่าวออนไลน์ที่ให้บริการข่าวด้านความปลอดภัย ได้แก่ เว็บสนุกดอทคอม (<http://news.sanook.com/rss/>) และเว็บ นสพ.คอม (<http://www.norsorpor.com/>) ซึ่งเว็บสนุกดอทคอม เป็นเว็บไซต์ที่ให้บริการข่าว (สนุก! ข่าว) ที่ได้รับความนิยมเป็นอันดับหนึ่งในช่วงที่มีเหตุการณ์ทางการเมืองอย่างเข้มข้น และเว็บ นสพ.คอม เป็นเว็บไซต์ที่รวบรวมข่าวประจำวันจากหนังสือพิมพ์ออนไลน์หลาย ๆ ฉบับมาเก็บไว้ในเว็บเดียว ข่าวต่าง ๆ จะถูกอัปเดตทุก ๆ 60 นาที โดยโปรแกรมอัตโนมัติที่ เว็บ นสพ.คอม ได้พัฒนาขึ้น ทำให้ผู้อ่านสามารถอ่านและเปรียบเทียบข่าวต่าง ๆ ได้อย่างรวดเร็ว ไม่เสียเวลาในการเปิดอ่านทีละฉบับ ปัจจุบันเว็บ นสพ.คอม ได้รวบรวมข่าวจากหนังสือพิมพ์ชั้นนำทั้งไทยและต่างประเทศ อาทิ ไทยรัฐ เคลิวิสต์ คมชัดลึก ผู้จัดการ มติชน ข่าวสด เดอะเนชั่น (The Nation) แอลจาซีรา (Aljazeera) และซีเอ็นเอ็น (CNN) เป็นต้น ทั้งนี้เพื่อให้ผู้ใช้สามารถเลือกดูข่าวสารที่ต้องการได้ง่ายและสะดวกรวดเร็วขึ้น

จากเว็บรวบรวมข่าวสนุกดอทคอมและนสพ.คอม พบว่า เว็บดังกล่าวมีหัวข้อข่าวที่เกี่ยวข้องกับความไม่ปลอดภัยที่หลากหลาย โดยเฉพาะเว็บ นสพ.คอม และจากการศึกษาวรรณกรรมที่เกี่ยวข้องกับความปลอดภัย (George, 2003; Prideaux, 2005; Rittichaiuwat and Chakraborty, 2008; สิรินาถ นุชย์เหล็ก, 2552; Howard, 2009; ปิ่นปิลันธน์ วังซ้าย, 2553; สมาคมนักวิชาการการท่องเที่ยว, 2553) สามารถแบ่งข่าวที่เกี่ยวข้องกับความปลอดภัยออกเป็น 5 ด้าน ได้แก่ ด้านสุขภาพ ด้านการจราจร ด้านภัยธรรมชาติ ด้านความไม่สงบของบ้านเมือง และด้านอาชญากรรม โดยแต่ละด้านจะได้มาจากข่าวแต่ละประเภทที่แตกต่างกัน ดังแสดงรายละเอียดในตารางที่ 1

ตารางที่ 1 การรวบรวมข่าวที่เกี่ยวข้องกับความไม่ปลอดภัยในแต่ละด้าน

ประเภทของความไม่ปลอดภัย	ข่าวที่เกี่ยวข้องกับความไม่ปลอดภัย	จำนวน (ข่าว)
ด้านสุขภาพ	ข่าวสุขภาพ	30
ด้านภัยธรรมชาติ	ข่าวการเกิดโรคระบาด	
	ข่าวน้ำท่วม	30
	ข่าวสึนามิ	
	ข่าวไฟไหม้ป่า	
ด้านความไม่สงบของบ้านเมือง	ข่าวสลายม็อบ	30
	ข่าวการเมือง	
	ข่าวการชุมนุม	
	ข่าวอุบัติเหตุ	30
ด้านการจราจร	ข่าวประจำวัน	
ด้านอาชญากรรม	ข่าวอาชญากรรม	30

1.2 วิเคราะห์คำศัพท์ที่เกี่ยวข้องกับความไม่ปลอดภัยจากข่าว โดยการวิเคราะห์คำศัพท์ที่เกี่ยวข้องกับความไม่ปลอดภัยนั้น จะเป็นการวิเคราะห์จากข่าวที่ได้รวบรวมมาจากขั้นตอนที่ 1.1 ซึ่งแบ่งออกเป็น 5 ด้าน ตามข่าวประเภทต่าง ๆ และในการวิเคราะห์คำศัพท์นั้น ทำโดยการจัดทำแบบสอบถามขึ้นเพื่อเป็นแนวทางให้กับผู้เชี่ยวชาญจำนวน 3 ท่าน ในการวิเคราะห์คำศัพท์ ดังตารางที่ 2 ซึ่งถ้าคำศัพท์ใดผู้เชี่ยวชาญเห็นด้วยว่าเป็นคำศัพท์ที่เกี่ยวข้องกับความไม่ปลอดภัยในข่าวประเภทนั้น ๆ จะให้คะแนนเป็น 1 ถ้าไม่เห็นด้วย จะให้คะแนนเป็น -1 แต่ถ้าไม่แน่ใจ จะให้คะแนนเป็น 0 และถ้าคะแนนเฉลี่ยของทั้ง 3 คน มากกว่า 0.5 คำศัพท์นั้นจะจัดเป็นคำศัพท์สำหรับข่าวประเภทนั้น ๆ แต่ถ้าต่ำกว่าจะไม่จัดว่าเป็นคำศัพท์สำหรับข่าวประเภทนั้น ๆ

ในการเลือกคำศัพท์ที่เกี่ยวข้องกับความไม่ปลอดภัยเพื่อเป็นแนวทางให้กับผู้เชี่ยวชาญ วิเคราะห์คำศัพท์นั้น ทำโดยการนำข่าวที่เป็นชุดข้อมูลศึกษา (Training set) จำนวน 150 ข่าว ที่ได้จากการรวบรวมข่าวในกระบวนการที่ 1.1 มาวิเคราะห์คำศัพท์ที่เกี่ยวข้องกับความไม่ปลอดภัยจากข่าวแต่ละด้าน และนำคำศัพท์ที่วิเคราะห์ได้นั้นไปสอบถามผู้เชี่ยวชาญ

ตารางที่ 2 ตัวอย่างแบบสอบถามในการวิเคราะห์คำศัพท์ที่เกี่ยวข้องกับความไม่ปลอดภัยของชาวประมงต่าง ๆ

ประเภทชาว	คำศัพท์ที่เกี่ยวข้องกับ ความไม่ปลอดภัย	คะแนนการพิจารณา			คะแนนเฉลี่ย
		ท่านที่ 1	ท่านที่ 2	ท่านที่ 3	
1. ด้านความไม่สงบของบ้านเมือง	1. ยุติการชุมนุม	1	1	1	1.00
	2. ยอมเปิดเส้นทาง	1	0	1	0.67
	3. ยอมสลายการชุมนุม	1	1	1	1.00
		...	...	...	...
2. ด้านอาชญากรรม	1. จับกุมผู้ต้องหา	1	1	1	1.00
	2. กองปราบปรามจับกุม	1	0	1	0.67
		...	...	...	...
		...	...	...	...

จากการเก็บข้อมูลในแบบสอบถาม สามารถสรุปคำศัพท์ที่เกี่ยวข้องกับความไม่ปลอดภัยในแต่ละด้านได้ดังต่อไปนี้

1) คำศัพท์ด้านสุขภาพ

ชาวความไม่ปลอดภัยด้านสุขภาพ มีคำศัพท์ที่เกี่ยวข้องกับเหตุการณ์ดังนี้

- โรคติดต่อ โรคไม่ติดต่อ ระวังภัยสุขภาพ สารเคมีตกค้าง สารเคมีอันตราย เป็นเชื้อที่มีอันตราย อาการระคายเคือง เป็นผื่นแดง ปวดแสบปวดร้อน เข้าสู่ร่างกายโดยทางเดินหายใจ เกิดอาการแพ้ ติดเชื้อโรคในกรณีรุนแรง อันตรายต่ออวัยวะภายใน อาจช็อกตาย เสี่ยงต่อการรับฝุ่นละออง ผู้ป่วยหายใจลำบาก ติดเชื้อไวรัส ร้ายแรงอาจทำให้เสียชีวิต ทำลายอวัยวะภายใน เชื้อแพร่กระจาย เกิดการระบาด ติดเชื้อเฉียบพลัน เตือนภัยเรื่องสุขภาพ ติดเชื้อในกระแสเลือด เขตควบคุมโรค สารพิษตกค้าง ติดต่อกับการไอ จาม เข้าสู่ร่างกายทางการหายใจ มีอาการหายใจผิดปกติ เสี่ยงต่อการติดเชื้อ ปนเปื้อนเชื้อไวรัส เชื้ออยู่ในอากาศ เกิดภาวะแทรกซ้อน อาจเกิดโรคแทรกซ้อน ผื่นขึ้นกระจายทั่วตัว ปวดกล้ามเนื้อ

2) คำศัพท์ด้านการจราจร

ชาวความไม่ปลอดภัยด้านการจราจร มีคำศัพท์ที่เกี่ยวข้องกับเหตุการณ์ดังนี้

- เกิดอุบัติเหตุรถชนกัน รถยนต์พลิกคว่ำ ประสานงาน ไม่สามารถส่งผู้บาดเจ็บไปหาการรักษาติดขัด รถยนต์ชนกับรถบรรทุก อุบัติเหตุรุนแรง บาดเจ็บสาหัส มีผู้เสียชีวิต เกิดไฟลุกไหม้ ประสบอุบัติเหตุ ถนนทรุด กีดขวางช่องทางการจราจร รถยนต์พลิกคว่ำ รถบรรทุกพลิกคว่ำ รถทัวร์พลิกคว่ำ การจราจรคับคั่ง อัดก๊อปปี้ รถแหกโค้ง รถสิบล้อพลิกคว่ำ บาดเจ็บเล็กน้อย รถไฟชนกับรถยนต์ สะพานถล่ม ชนราวสะพาน ร่วงตกเหว ร่วงตกสะพาน เกิดอุบัติเหตุรถพุ่ง 18 ล้อ

3) คำศัพท์ด้านภัยธรรมชาติ

ข่าวความปลอดภัยด้านภัยธรรมชาติมีคำศัพท์ที่เกี่ยวข้องกับเหตุการณ์ดังนี้

- น้ำป่าไหลหลาก ฝนฟ้าคะนอง น้ำป่าพัด ฝนตกหนัก ระดับน้ำอยู่ในระดับวิกฤติ น้ำไหลบ่า ฝนตกชุกหนาแน่น เข้าขั้นวิกฤติ ดินถล่ม ถนนขาด ฝนตกติดต่อกันหลายวัน น้ำท่วม ถูกตัดขาดจากโลกภายนอก อุทกภัย ถูกตัดขาดอย่างสิ้นเชิง ฝนตกอย่างต่อเนื่อง ถนนทรุด โคลนถล่ม ทะลักเข้าท่วม ระดับน้ำท่วมสูง มรสุม พายุ น้ำท่วม น้ำทะเลหนุน น้ำเหนือไหลมาสมทบ ระดับน้ำไม่ถึงระดับวิกฤติ น้ำท่วมฉับพลัน อพยพ

4) คำศัพท์ด้านความไม่สงบของบ้านเมือง

ข่าวความปลอดภัยด้านความไม่สงบของบ้านเมืองมีคำศัพท์ที่เกี่ยวข้องกับเหตุการณ์ดังนี้

- ก่อม็อบ สร้างความไม่สงบ ก่อการร้าย กีดขวางการจราจร กลุ่มผู้ชุมนุมก่อความไม่สงบ ทำลายข้าวของปิดถนน มีผู้บาดเจ็บ ยุติการชุมนุม ประท้วง สลายการชุมนุม ประกาศเคอร์ฟิว ขวางป่าข้าวของ เหตุการณ์ไม่ปกติ จุดไฟฟ้า จอรถขวางถนน ขวางการจราจร ผู้ชุมนุมเคลื่อนไหว มีผู้เสียชีวิต มีการปะทะกัน ใช้แก๊สน้ำตา ใช้น้ำฉีด

5) คำศัพท์ด้านอาชญากรรม

ข่าวความปลอดภัยด้านอาชญากรรมมีคำศัพท์ที่เกี่ยวข้องกับเหตุการณ์ดังนี้

- ถูกโจรปล้น ฆ่าชิงทรัพย์ มีโจรชิงทรัพย์ จับกุมผู้ต้องหา เกิดการยิงประทะ เกิดเหตุฆาตกรรม เกิดเหตุระเบิด พบวัตถุระเบิด ชิงทรัพย์ ถูกวิสามัญ เกิดการปะทะกัน ก่อเหตุสะเทือนขวัญ มีผู้เสียชีวิต พบศพ ลักทรัพย์ ผู้บาดเจ็บ ก่อการร้าย ถูกคนร้ายฆ่า กองปราบปรามจับกุม กระทำชำเรา มีอาวุธปืน กระทำทารุณ

1.3 วิเคราะห์ระดับความรุนแรงของเหตุการณ์จากคำศัพท์ ในการวิเคราะห์ระดับความรุนแรงของเหตุการณ์นั้น ได้มีการจัดทำแบบสอบถามขึ้นเพื่อเป็นแนวทางในการจัดระดับความรุนแรงของเหตุการณ์ โดยมีการสอบถามจากผู้เชี่ยวชาญจำนวน 30 คน และได้ข้อมูลดังแสดงตัวอย่างในตารางที่ 3 ซึ่งเป็นข้อมูลที่ได้จากการสอบถามความคิดเห็นเกี่ยวกับระดับความรุนแรงเมื่อพบคำศัพท์ที่เกี่ยวข้องกับภัยธรรมชาติต่าง ๆ โดยสามารถสรุประดับความรุนแรงของเหตุการณ์จากคำศัพท์ได้ดังตารางที่ 4

ตารางที่ 3 ตัวอย่างข้อมูลที่ได้จากแบบสอบถามระดับความรุนแรงจากคำศัพท์ด้านภัยธรรมชาติ

คำศัพท์ที่เกี่ยวข้องกับ ความปลอดภัยด้านภัยธรรมชาติ	ความคิดเห็น (โปรดระบุระดับความรุนแรงที่ท่านคิดเห็น เมื่อพบคำศัพท์ที่เกี่ยวข้องต่าง ๆ) 1 = ปลอดภัย 2 = ไม่เลวร้าย 3 = ค่อนข้างเลวร้าย 4 = เลวร้าย			
	1	2	3	4
ไม่มีฝนตก	26	3	0	1
น้ำลดลงหมดแล้ว	25	3	2	0
มีฝน/ฝนฟ้าคะนอง	2	27	1	0
น้ำทะเลหนุน	2	24	4	0
ทะเลลึกเข้าท่วม	0	1	24	5
ระดับน้ำอยู่ในระดับวิกฤต	0	10	12	8
น้ำป่าไหลหลาก/น้ำป่าพัด/น้ำไหลบ่า	0	0	1	29
ดินถล่ม	0	0	1	29
ถนนขาด	0	0	1	29

ตารางที่ 4 ตัวอย่างคำศัพท์ด้านภัยธรรมชาติที่มีระดับความรุนแรงต่างๆ

ระดับความรุนแรง	คำศัพท์ด้านภัยธรรมชาติ
ระดับที่ 1 ปลอดภัย	ไม่มีฝนตก น้ำลดลงหมดแล้ว
ระดับที่ 2 มีเหตุการณ์ แต่ไม่เลวร้าย	มีฝน ฝนฟ้าคะนอง น้ำทะเลหนุน
ระดับที่ 3 มีเหตุการณ์ ค่อนข้างเลวร้าย	ทะเลลึกเข้าท่วม ระดับน้ำอยู่ในระดับวิกฤต
ระดับที่ 4 มีเหตุการณ์ เลวร้าย	น้ำป่าไหลหลาก น้ำป่าพัด น้ำไหลบ่า ดินถล่ม ถนนขาด

1.4 ทำนายระยะเวลาสิ้นสุดของเหตุการณ์จากคำศัพท์ ซึ่งเหตุที่ต้องมีการทำนายระยะเวลาสิ้นสุดของเหตุการณ์ เนื่องจากข่าวแต่ละด้านจะมีช่วงเวลาของการเกิดความรุนแรงที่ต่างกัน และการนำเสนอข่าวส่วนใหญ่จะมีเฉพาะขณะเกิดเหตุการณ์ในช่วงแรกเท่านั้น แม้ว่าเหตุการณ์ยังไม่สิ้นสุด แต่ก็ไม่มีการนำเสนอข่าวแล้ว เช่น ข่าวด้านสุขภาพ มีการนำเสนอข่าวการระบาดของไข้หวัดนกในโคราช เมื่อวันที่ 1 มกราคม 2554 และพบว่าการระบาดโดยปกติจะมีระยะเวลา 3 เดือนจึงจะสิ้นสุด ดังนั้นเมื่อนักท่องเที่ยวต้องการเดินทางวันที่ 3 กุมภาพันธ์ 2554 แม้ว่าจะไม่มีข่าวที่นำเสนอเกี่ยวกับการระบาดของไข้หวัดนก ณ วันนั้น แต่การระบาดยังคงมีอยู่

ดังนั้นวิธีการสกัดข่าวด้านความปลอดภัย เพื่อนำเสนอเส้นทางที่เหมาะสมให้กับนักท่องเที่ยว จึงจำเป็นต้องมีการทำนายระยะเวลาสิ้นสุดของเหตุการณ์ความรุนแรงที่เกิดขึ้น เพื่อให้สามารถแจ้งเตือนให้กับนักท่องเที่ยว เพื่อหลีกเลี่ยงการเดินทางไปโคราชได้

ซึ่งจากการวิเคราะห์ในเบื้องต้นพบว่า การทำนายระยะเวลาสิ้นสุดของเหตุการณ์ความไม่ปลอดภัยขึ้นอยู่กับ วันที่ในการนำเสนอข่าว ประเภทของเหตุการณ์ที่เกิดขึ้น และระดับความรุนแรงของเหตุการณ์ที่ต่างกัน ดังนั้นการทำนายระยะเวลาสิ้นสุดของเหตุการณ์นั้น ต้องอาศัยข้อมูลเหล่านี้ รวมทั้งในการวิจัยนี้ จะอาศัยข้อมูลจากผู้เชี่ยวชาญเพื่อตรวจสอบความเหมาะสมในการทำนาย โดยในตารางที่ 5 แสดงตัวอย่างของการเก็บข้อมูลความคิดเห็นของผู้เชี่ยวชาญจำนวน 30 คนจากแบบสอบถามที่ใช้ในการทำนายระยะเวลาสิ้นสุดของเหตุการณ์ด้านภัยธรรมชาติ

ตารางที่ 5 ตัวอย่างแบบสอบถามการทำนายระยะเวลาสิ้นสุดของเหตุการณ์ด้านภัยธรรมชาติ

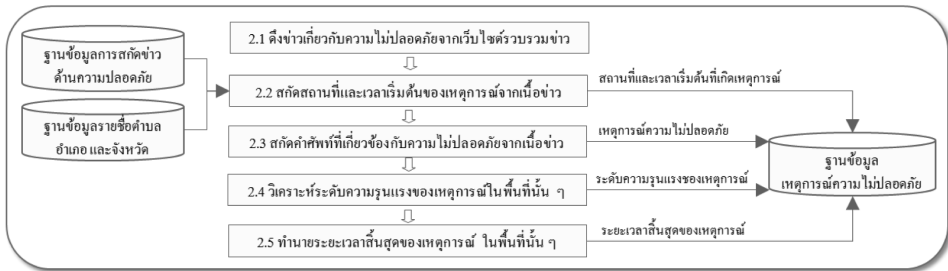
คำศัพท์ที่เกี่ยวข้องกับ ความปลอดภัยด้านการจราจร	ความคิดเห็น (หากท่านไม่เห็นด้วยกับช่วงเวลาที่ระบุไว้โปรดระบุช่วงเวลาที่คุณคิดว่าควรจะสิ้นสุด เหตุการณ์ที่ท่านคิดเห็น)					
	0-3 วัน	1 สัปดาห์	2 สัปดาห์	3 สัปดาห์	4 สัปดาห์	อื่นๆ ระบุ
ไม่มีฝนตก	25	4	1	0	0	
มีฝน/ฝนฟ้าคะนอง	26	4	0	0	0	
น้ำทะเลหนุน	25	5	0	0	0	
ทะเลลึกเข้าท่วม	3	0	26	0	1	
ระดับน้ำอยู่ในระดับวิกฤต	4	23	3	0	0	
น้ำป่าไหลหลาก/น้ำป่าพัด/น้ำไหลบ่า	5	16	7	2	0	
ดินถล่ม	1	6	2	1	20	
ถนนขาด	1	6	2	1	20	

จากข้อมูลที่ได้จากการตอบแบบสอบถามนั้น พบว่า ระยะเวลาสิ้นสุดของเหตุการณ์มีหน่วยนับที่แตกต่างกัน 3 หน่วย คือ ชั่วโมง วัน และ สัปดาห์ซึ่งหน่วยนับขึ้นอยู่กับประเภทของเหตุการณ์และถึงแม้ว่าคำศัพท์บางคำจะมีระดับความรุนแรงเท่ากัน แต่อาจมีระยะเวลาสิ้นสุดของเหตุการณ์ต่างกันได้ ซึ่งสามารถสรุประยะเวลาสิ้นสุดของเหตุการณ์จากคำศัพท์ได้ดังตารางที่ 6

ตารางที่ 6 ตัวอย่างคำศัพท์ด้านภัยธรรมชาติที่มีระยะเวลาการสิ้นสุดของเหตุการณ์ต่างกัน

ระยะเวลาสิ้นสุดเหตุการณ์	คำศัพท์ด้านภัยธรรมชาติ
ระยะเวลา 0-3 วัน	ไม่มีฝนตก มีฝน ฝนฟ้าคะนอง น้ำทะเลหนุน
ระยะเวลา 1 สัปดาห์	ระดับน้ำอยู่ในระดับวิกฤต น้ำป่าไหลหลาก น้ำป่าพัด น้ำไหลบ่า
ระยะเวลา 2 สัปดาห์	ทะเลลึกเข้าท่วม
ระยะเวลา 3 สัปดาห์	-
ระยะเวลา 4 สัปดาห์	ดินถล่ม ถนนขาด

ขั้นตอนที่ 2 สกัดข่าวหรือเหตุการณ์ความไม่ปลอดภัยที่เกิดขึ้นในช่วงเวลานั้นจากเว็บไซต์รวบรวมข่าว โดยการสกัดข่าวหรือเหตุการณ์ความไม่ปลอดภัยที่เกิดขึ้นมีการการดำเนินงานดังภาพที่ 3



ภาพที่ 3 ขั้นตอนสกัดข่าวหรือเหตุการณ์ความไม่ปลอดภัยที่เกิดขึ้นในช่วงเวลานั้นจากเว็บไซต์รวบรวมข่าว

2.1 ดึงข่าวเกี่ยวกับความไม่ปลอดภัยจากเว็บไซต์รวบรวมข่าว โดยการดึงข่าวที่เกี่ยวข้องกับความไม่ปลอดภัยนั้นได้ทำการเขียนโปรแกรมเพื่อดึงข่าวแบบอัตโนมัติจากเว็บไซต์รวบรวมข่าว (นสพ.คอม และข่าวสนุกคอทคอม) โดยเลือกเฉพาะข่าวด้านความปลอดภัยและแยกประเภทของข่าวตามที่กล่าวมาข้างต้น เช่น ดึงข่าวด้านภัยธรรมชาติจากเว็บไซต์จะได้ข่าวน้ำท่วม ดังภาพที่ 4 หลังจากนั้นจะนำข่าวที่มีอยู่ไปใช้ในการสกัดสารสนเทศต่อไป

2.2 สกัดสถานที่และเวลาเริ่มต้นของเหตุการณ์จากเนื้อหาข่าว โดยใช้หลักการทำงานของ การสกัดสารสนเทศจากพจนานุกรมคำศัพท์ ซึ่งคำศัพท์ที่ใช้ในการสกัดนั้นได้มาจากฐานข้อมูลรายชื่อตำบล อำเภอ และจังหวัดในประเทศไทย ฐานข้อมูลดังกล่าวจะนำมาช่วยในการระบุพื้นที่และขอบเขตของเหตุการณ์ที่พบในเนื้อหาข่าวซึ่งจากตัวอย่างของข่าวเกี่ยวกับน้ำท่วมจากเว็บ นสพ.คอม ที่รวบรวมข่าวจากหนังสือพิมพ์ข่าวสดในภาพที่ 4 สามารถระบุพื้นที่ ขอบเขตของเหตุการณ์ได้คือ พื้นที่ ต.ท่าไม้ อ.ชุมแสง จ.นครสวรรค์ ซึ่งข้อมูลดังกล่าวเป็นข้อมูลที่ตรงกับข้อมูลในฐานข้อมูลของพื้นที่ตำบล อำเภอ และจังหวัด ที่มีในประเทศไทย ระบบจึงสามารถสกัดออกมาได้ เพราะในการรายงานข่าวบางครั้งอาจไม่ระบุคำว่า จังหวัด หรือ จ. นำหน้าชื่อจังหวัดก็เป็นได้

ส่วนการสกัดวันเวลาเริ่มต้นของเหตุการณ์จะสกัดจากเนื้อหาข่าวก่อนหากในเนื้อหาข่าวไม่มีระบุวันที่ ก็จะใช้วันที่ประกาศข่าวหรือวันที่เผยแพร่ข่าวแทน ซึ่งในตัวอย่างนี้จะสกัดได้จากวันที่ประกาศข่าวหรือวันที่เผยแพร่ข่าว คือวันที่ 1 ธันวาคม 2554



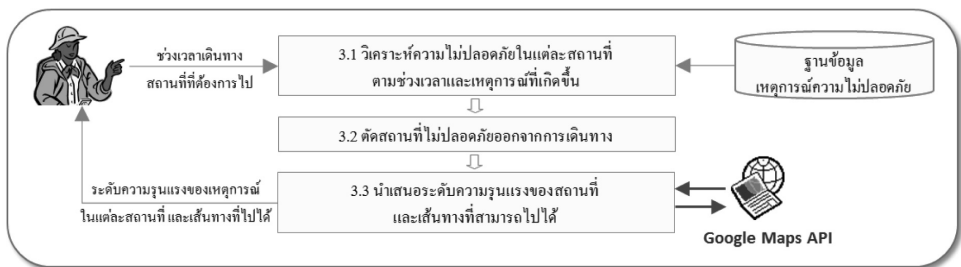
อย่างไรก็ตามในการสกัดคำศัพท์ที่ได้จากข่าวใดข่าวหนึ่งนั้น อาจได้คำศัพท์ที่เกี่ยวข้องกับ

ระดับความรุนแรง มากกว่า 1 ระดับ เช่น ข่าวด่วนว่า “ในพื้นที่ภาคใต้ มีฝนตกหนัก และมีดินถล่มเกิดขึ้น” ซึ่งจากข่าวจะได้รับความรุนแรงทั้งในระดับที่ 2 (จากคำศัพท์ ฝนตกหนัก) และระดับที่ 4 (จากคำศัพท์ ดินถล่ม) ซึ่งจากการศึกษา และวิเคราะห์ข่าวต่าง ๆ ที่เกิดขึ้น และจากการสอบถามผู้เชี่ยวชาญในเบื้องต้น สามารถวินิจฉัยข่าวได้ว่า มีเหตุการณ์ความรุนแรงในระดับที่ 4 นั่นก็คือ จะจัดระดับความรุนแรงของเหตุการณ์ ตามคำศัพท์ที่บ่งบอกระดับความรุนแรงที่มากที่สุดที่พบในข่าวนั้น ๆ

2.5 ทำนายระยะเวลาสิ้นสุดของเหตุการณ์ในพื้นที่นั้น ๆ ในการทำนายระยะเวลาสิ้นสุดของเหตุการณ์ จะทำนายระยะเวลาสิ้นสุดของเหตุการณ์จากคำศัพท์ในลักษณะเดียวกับการวิเคราะห์ระดับความรุนแรงของเหตุการณ์ คือใช้ข้อมูลที่ได้ขั้นตอนที่ 1.4

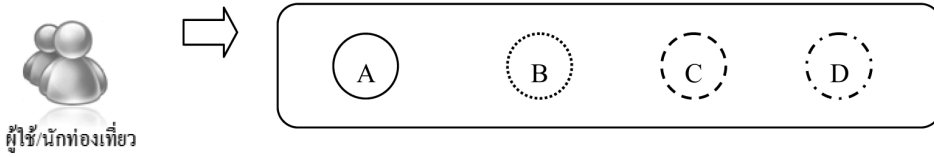
ซึ่งจากตัวอย่างของข่าวเกี่ยวกับน้ำท่วมในภาพที่ 4 คำศัพท์ที่เกี่ยวข้องกับความไม่ปลอดภัยได้คือคำว่า “น้ำท่วม” เมื่อนำมาทำนายระยะเวลาสิ้นสุดของเหตุการณ์จะพบว่าเหตุการณ์น้ำท่วมนั้นจะสิ้นสุดไม่เกิน 2 สัปดาห์ หรือ 14 วัน ซึ่งจะนำมารวมกับเวลาเริ่มต้นของเหตุการณ์ ก็จะได้เวลาสิ้นสุดของเหตุการณ์ ซึ่งจากตัวอย่างจะได้เวลาสิ้นสุดของเหตุการณ์คือ วันที่ 15 ธันวาคม 2554 (มาจากวันเริ่มต้นเหตุการณ์ คือ 1 ธันวาคม 2554 + 14 วัน)

ขั้นตอนที่ 3 วิเคราะห์ความไม่ปลอดภัยที่เกิดขึ้นของสถานที่ที่ต้องการไป โดยนักท่องเที่ยวจะระบุสถานที่ต่าง ๆ ที่ต้องการไป และช่วงเวลาเดินทาง ขั้นตอนนี้จะวิเคราะห์หาระดับความรุนแรงของสถานที่เหล่านั้นตามช่วงเวลาที่ต้องการเดินทาง ซึ่งถ้า ณ เวลานั้น พบเหตุการณ์ความไม่ปลอดภัยที่ร้ายแรงเกิดขึ้น ก็จะตัดสถานที่นั้นทิ้ง หรือไม่ให้นักท่องเที่ยวเดินทางไป โดยมีการดำเนินการแสดงดังภาพที่ 5



ภาพที่ 5 ขั้นตอนวิเคราะห์ความไม่ปลอดภัยที่เกิดขึ้นของสถานที่ที่ต้องการไป

3.1 วิเคราะห์ความไม่ปลอดภัยในแต่ละสถานที่ตามช่วงเวลาและเหตุการณ์ที่เกิดขึ้น จากการวิเคราะห์ระดับความรุนแรงของเหตุการณ์ตามคำศัพท์ที่สกัดได้ทั้งหมดในขั้นตอนที่ 2.4 จะทำให้ทราบสถานการณ์ความไม่ปลอดภัยในแต่ละสถานที่ โดยการวิเคราะห์ความไม่ปลอดภัยในแต่ละสถานที่ที่จะยึดเอาระดับความรุนแรงสูงสุดของคำศัพท์ที่สกัดได้โดยภาพที่ 6 แสดงตัวอย่างการวิเคราะห์ระดับความรุนแรง เมื่อข้อมูลนำเข้า คือสถานที่ที่ต้องการไป 4 แห่ง



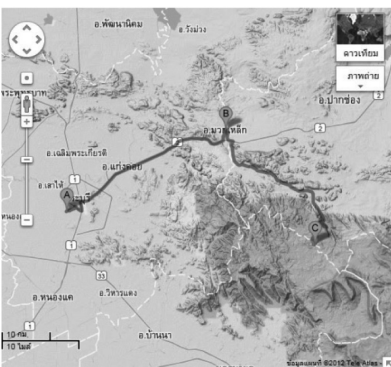
ภาพที่ 6 ตัวอย่างการวิเคราะห์ความไม่ปลอดภัยในแต่ละสถานที่ตามช่วงเวลาและเหตุการณ์ที่เกิดขึ้น

จากรูปเป็นตัวอย่างสถานที่ที่มีความรุนแรงแตกต่างกัน 4 ระดับ โดยรูปแบบของเส้นมีลักษณะดังนี้

- เส้นทึบ (—) หมายถึง สถานที่ที่ปลอดภัยไม่มีเหตุการณ์สามารถเดินทางได้
- เส้นจุดสี่เหลี่ยมเล็ก (.....) หมายถึง สถานที่ที่เกิดเหตุการณ์แต่ไม่รุนแรงสามารถเดินทางได้
- เส้นปะขนาดกลางสม่ำเสมอ (_ _) หมายถึง สถานที่ที่เกิดเหตุการณ์ค่อนข้างรุนแรงสามารถเดินทางแต่ควรที่จะหลีกเลี่ยงเส้นทาง
- เส้นปะและจุดสี่เหลี่ยม (_ .) หมายถึง สถานที่ที่เกิดเหตุการณ์ที่รุนแรงมาก จนไม่สามารถเดินทางไปได้

3.2 คัดสถานที่ไม่ปลอดภัยออกจากการเดินทาง เมื่อได้ระยะเวลาสิ้นสุดของเหตุการณ์ ความรุนแรงของเหตุการณ์ และพื้นที่ขอบเขตของเหตุการณ์แล้ว สามารถนำข้อมูลเหล่านั้นมาเป็นข้อมูลที่ใช้ในการตัดสินใจสำหรับคัดสถานที่ไม่ปลอดภัยออกจากการเดินทาง ในภาพที่ 6 สถานที่ที่ D จะถูกตัดทิ้งออกจากแผนการเดินทางทันที

3.3 นำเสนอระดับความรุนแรงของสถานที่และเส้นทางที่สามารถไปได้ ในการนำเสนอระดับความรุนแรงของเหตุการณ์ในแต่ละสถานที่ที่สามารถไปได้พร้อมขอเส้นทางระหว่างสถานที่และแผนที่ภูมิประเทศจากส่วนต่อประสานกับผู้ใช้ของแผนที่กูเกิ้ล (Google Maps API) เพื่อช่วยในการสนับสนุนการตัดสินใจเลือกสถานที่ของนักท่องเที่ยวดังแสดงในภาพที่ 7



สถานที่ A ปลอดภัยไม่มีเหตุการณ์ สามารถเดินทางได้

สถานที่ B เกิดเหตุการณ์แต่ไม่รุนแรง สามารถเดินทางได้

สถานที่ C เกิดเหตุการณ์ค่อนข้างรุนแรง ควรหลีกเลี่ยง

ภาพที่ 7 สถานที่ และเส้นทางที่ได้จากขั้นตอนการตัดเส้นทางที่ไม่ปลอดภัยออกจากการเดินทางท่องเที่ยว

จากภาพที่ 7 จะแสดงสถานที่และเส้นทางที่ได้จากขั้นตอนที่ 3.2 ที่ตัดสถานที่ที่ไม่ปลอดภัยออกจากแผนการเดินทางท่องเที่ยวแล้ว พร้อมคำอธิบายระดับความรุนแรงของเหตุการณ์ในแต่ละสถานที่ เพื่อช่วยในการตัดสินใจในการเดินทางให้กับนักท่องเที่ยว โดยหลังจากที่นักท่องเที่ยวได้ระบุสถานที่ที่ต้องการไปท่องเที่ยวแล้ว ค่าคะแนนความไม่ปลอดภัยเหล่านี้ จะถูกนำไปคำนวณรวมกับสถิติความไม่ปลอดภัยในอดีต และส่วนสนับสนุนด้านความปลอดภัย เพื่อจัดลำดับแผนการท่องเที่ยวภายใต้เงื่อนไขบังคับด้านความปลอดภัย และนำเสนอให้กับนักท่องเที่ยวต่อไป

การประเมินวิธีการสกัดข่าว

การวิเคราะห์ข้อมูลเพื่อประเมินความถูกต้องของวิธีการสกัดข่าวที่นำเสนอ นั้น จะทำโดยการนำวิธีการที่ได้ไปสกัดข่าวจริงที่นำมาจากเว็บรวบรวมข่าวออนไลน์ ได้แก่ เว็บสนุกดอทคอม และเว็บ นสพ.คอม โดยมีรายละเอียดของการประเมินความถูกต้องดังนี้

1. นำวิธีการสกัดข่าวไปใช้ในการสกัดข่าวออนไลน์จริง โดยข่าวที่นำมาใช้เป็นชุดข้อมูลทดสอบ (Testset) จำนวน 150 ข่าว (5 ด้าน ด้านละ 30 ข่าว) เป็นคนละชุดกับข่าวที่นำมาใช้ในการพัฒนาวิธีการ (Training Set)

2. ตรวจสอบความถูกต้องของข่าว โดยการวิเคราะห์ข่าวที่มีอยู่ในชุดข้อมูลทดสอบทีละข่าว โดยผู้เชี่ยวชาญ ว่ามีคำศัพท์ที่เกี่ยวข้องกับความไม่ปลอดภัยในด้านนั้น ๆ อยู่จริงกี่คำ และเหตุการณ์มีระดับความรุนแรงในระดับใด และทำการเก็บรวบรวม นับจำนวนคำศัพท์ที่พบ และระดับความรุนแรงไว้ เพื่อนำไปใช้ในการประเมินในส่วนต่อไป

3. ประเมินความถูกต้องของวิธีการสกัดคำศัพท์ที่เกี่ยวข้องกับความไม่ปลอดภัย โดยคิดเป็นร้อยละของค่าความสัมพันธ์ระหว่าง จำนวนคำศัพท์ที่สกัดได้จากขั้นตอนวิธีการในการสกัดข่าว และจำนวนคำศัพท์ด้านความไม่ปลอดภัยทั้งหมดที่มีในข่าวที่เป็นชุดทดสอบ ดังสมการที่ 1

$$\text{ร้อยละความถูกต้องในการสกัดคำศัพท์} = \frac{\text{จำนวนคำศัพท์ที่สกัดได้จากขั้นตอนวิธี}}{\text{จำนวนคำศัพท์ด้านความไม่ปลอดภัยทั้งหมดที่มี}} \times 100 \quad (1)$$

4. ประเมินความคลาดเคลื่อนของวิธีการวิเคราะห์ระดับความรุนแรงของเหตุการณ์โดยพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error: MAE) เป็นตัวแปรที่แสดงถึงความคลาดเคลื่อนโดยไม่คำนึงถึงทิศทางบวกหรือลบ ระหว่างค่าที่ได้จากการวิเคราะห์ของวิธีการ กับค่าของข้อมูลจริงในชุดทดสอบ ซึ่งค่าที่ได้หากมีค่าเข้าใกล้ศูนย์ ความแม่นยำในการวิเคราะห์ของวิธีการจะสูงตามไปด้วย โดยค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (MAE) แสดงดังสมการที่ 2

$$\text{ความคลาดเคลื่อนในการวิเคราะห์ระดับความรุนแรง} = \frac{1}{N} \sum_{i=1}^N |Q_i - F_i| \quad (2)$$

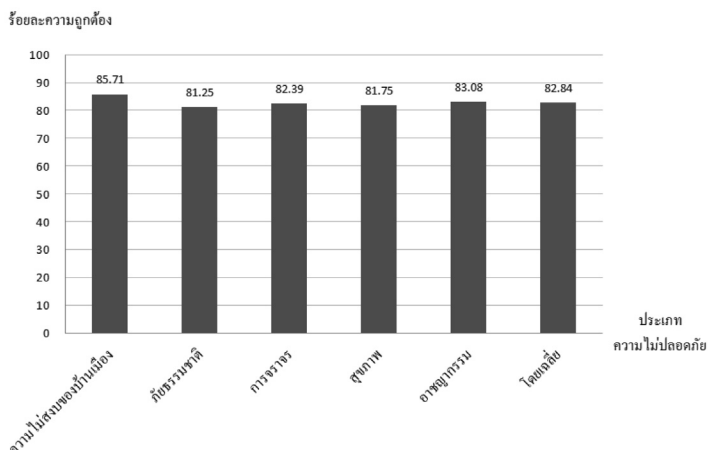
โดยที่ Q_i = ระดับความรุนแรงที่ได้จากวิเคราะห์ของวิธีการ; F_i = ระดับความรุนแรงจริงในชุดข้อมูลทดสอบ; i = ข่าวที่; และ N = จำนวนข่าวทั้งหมด

ผลการประเมิน และการอภิปรายผล

จากการนำวิธีการสกัดข่าวไปใช้ในการสกัดข่าวออนไลน์จริงกับข่าวที่ใช้เป็นชุดข้อมูลทดสอบ (Testset) มีผลการทดสอบและอภิปรายผลดังนี้

1. ผลการประเมินความถูกต้องในการสกัดคำศัพท์ที่เกี่ยวข้องกับความปลอดภัย โดยคิดเป็นร้อยละของค่าความสัมพันธ์ระหว่างจำนวนคำศัพท์ที่สกัดได้จากวิธีการสกัดข่าว และจำนวนคำศัพท์ด้านความปลอดภัยทั้งหมดที่มีในข่าวที่เป็นชุดทดสอบจำนวน 150 ข่าว (5 ด้าน ด้านละ 30 ข่าว) มีความถูกต้องในการสกัดข่าวด้านความไม่สงบของบ้านเมือง 85.71% ด้านภัยธรรมชาติ 81.25% ด้านการจราจร 82.39% ด้านสุขภาพ 81.75% ด้านอาชญากรรม 83.08% และมีความถูกต้องในการสกัดข่าวโดยเฉลี่ยทั้ง 5 ด้าน 82.84% แสดงดังภาพที่ 5

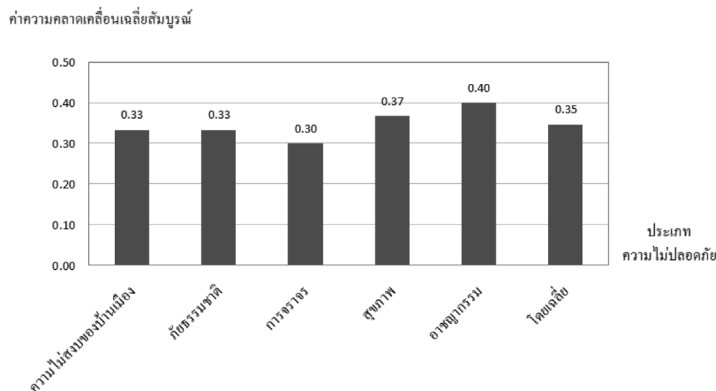
จากผลการประเมินความถูกต้องพบว่า การสกัดคำศัพท์ที่เกี่ยวข้องกับความปลอดภัยด้านความไม่สงบของบ้านเมืองมีค่าความถูกต้องสูงที่สุดคือ 85.71% ทั้งนี้เนื่องจากฐานข้อมูลคำศัพท์ที่เกี่ยวข้องกับปลอดภัยในด้านความไม่สงบของบ้านเมืองมีคำศัพท์ที่ให้ความหมายชัดเจนและเป็นคำศัพท์ที่มีคำศัพท์อื่น ๆ ที่มีคำที่คล้ายคลึงกับคำศัพท์และมีความเดียวกับคำศัพท์นั้น ๆ ไม่มากนักต่างจากคำศัพท์ที่เกี่ยวข้องกับความปลอดภัยด้านอื่น ๆ ที่มีความหมายชัดเจนแต่มีคำศัพท์ที่คล้ายคลึงกับคำศัพท์นั้นและมีความหมายเดียวกันกับคำศัพท์นั้น ๆ ซึ่งไม่สามารถเพิ่มคำศัพท์เหล่านั้นเข้าไปในฐานข้อมูลได้เนื่องจากมีคำเหล่านั้นจำนวนมาก อีกทั้งจะทำให้ฐานข้อมูลคำศัพท์ที่เกี่ยวข้องกับปลอดภัยในแต่นั้นมีขนาดใหญ่จนเกินไป ส่งผลให้การทำงานของวิธีการสกัดข่าวใช้เวลาในการทำงานนานมากขึ้นอีกด้วย



ภาพที่ 5 กราฟแสดงค่าร้อยละความถูกต้องในการสกัดคำศัพท์ที่เกี่ยวข้องกับความปลอดภัย

2. ผลการประเมินความคลาดเคลื่อนในการวิเคราะห์ระดับความรุนแรงของเหตุการณ์ โดยพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (*Mean Absolute Error: MAE*) เป็นตัวแปรที่แสดงถึงความคลาดเคลื่อนโดยไม่คำนึงถึงทิศทางบวกหรือลบ ระหว่างค่าที่ได้จากการวิเคราะห์ของวิธีการ กับค่าของข้อมูลจริงในชุดทดสอบ มีค่าความคลาดเคลื่อนในการวิเคราะห์ระดับความรุนแรงของเหตุการณ์ ด้านความไม่สงบของบ้านเมือง 0.33 ด้านภัยธรรมชาติ 0.33 ด้านการจราจร 0.30 ด้านสุขภาพ 0.37 ด้านอาชญากรรม 0.40 และมีค่าความคลาดเคลื่อนในการวิเคราะห์ระดับความรุนแรงของเหตุการณ์ โดยเฉลี่ยทั้ง 5 ด้านเท่ากับ 0.35 แสดงดังภาพที่ 6

จากผลการประเมินความคลาดเคลื่อนในการวิเคราะห์ระดับความรุนแรงของเหตุการณ์ พบว่าความคลาดเคลื่อนในการวิเคราะห์ระดับความรุนแรงของเหตุการณ์ด้านการจราจร มีค่าความคลาดเคลื่อนน้อยที่สุดคือ 0.3 ทั้งการวิเคราะห์ระดับความรุนแรงนั้นจะขึ้นอยู่กับเหตุการณ์ที่เกิดขึ้น ซึ่งแต่ละเหตุการณ์ก็จะมีระดับความรุนแรงที่แตกต่างกันออกไปตามเหตุการณ์ที่เกิดขึ้น ดังนั้นการวิเคราะห์ระดับความรุนแรงของเหตุการณ์จึงจำเป็นต้องอาศัยความคิดเห็นจากผู้เชี่ยวชาญดังที่ได้แสดงในแบบสอบถามดังที่กล่าวมาแล้ว



ภาพที่ 6 กราฟแสดงค่าความคลาดเคลื่อนในการวิเคราะห์ระดับความรุนแรงของเหตุการณ์

บทสรุป

งานวิจัยนี้ได้ออกแบบและพัฒนาวิธีการสกัดข่าวด้านความปลอดภัย เพื่อสนับสนุนระบบวางแผนการท่องเที่ยวออนไลน์ภายใต้เงื่อนไขบังคับด้านความปลอดภัย ให้สามารถวิเคราะห์และปรับปรุงข่าวสารด้านความปลอดภัยให้ทันสมัยอยู่เสมอ เพื่อให้ระบบวางแผนการท่องเที่ยวดังกล่าวมีประสิทธิภาพมากยิ่งขึ้น ผลการทดลองพบว่าวิธีการที่นำเสนอมีค่าความถูกต้องในการสกัดข่าวโดยเฉลี่ย 82.84% ประกอบด้วย ความถูกต้องในการสกัดข่าวด้านสุขภาพ 81.75% ด้านการจราจร 82.39% ด้านภัยธรรมชาติ 81.25% ด้านความไม่สงบของบ้านเมือง 85.71% ด้านอาชญากรรม 83.08% และมีค่าความคลาดเคลื่อนในการวิเคราะห์ระดับความรุนแรงของเหตุการณ์โดยเฉลี่ยเท่ากับ 0.35 ประกอบ

ด้วยความคลาดเคลื่อนในการวิเคราะห์ระดับความรุนแรงของเหตุการณ์ด้านสุขภาพ 0.37 ด้านการจราจร 0.30 ด้านภัยธรรมชาติ 0.33 ด้านความไม่สงบของบ้านเมือง 0.33 และด้านอาชญากรรม 0.40

ทั้งนี้การออกแบบและพัฒนาวิธีการสกัดข่าวด้านความปลอดภัย จะมีประสิทธิภาพในการสกัดมากขึ้น หากมีการนำเอาหลักการตัดคำโดยใช้เอ็นแกรมเข้ามาช่วยเนื่องจากในระบบการตัดคำด้วยพจนานุกรมมักพบปัญหาในกรณีที่คำที่ปรากฏในเอกสารนั้นไม่มีอยู่ในพจนานุกรม จึงจำเป็นต้องใช้เอ็นแกรมเข้ามาช่วยในการหาคำที่มีความหมายเดียวกันแต่อาจสะกดคำต่างกันเล็กน้อย เช่น มีคำว่า ฝนตกหนัก ในพจนานุกรม แต่ในเนื้อหาพบคำว่า ฝนตกอย่างหนัก ดังนั้นเมื่อใช้เอ็นแกรมเข้ามาช่วยจะทำให้สามารถพบคำที่ใกล้เคียงคำว่า ฝนตกอย่างหนัก มากที่สุด ก็คือ คำว่า ฝนตกหนัก ในพจนานุกรม ดังนั้นงานวิจัยในอนาคตจึงมุ่งเน้นการใช้เอ็นแกรมเข้ามาช่วยเพิ่มประสิทธิภาพของวิธีการสกัดข่าวด้านความปลอดภัยต่อไป

เอกสารอ้างอิง

- กระทรวงการท่องเที่ยวและกีฬา. (2553). วารสารกระทรวงการท่องเที่ยวและกีฬา. วารสารกระทรวงการท่องเที่ยวและกีฬา. 6(42): หน้า 9.
- นัฐวุฒิ ไชยเจริญ. (2544). การตัดคำและการกำกับหมวดคำภาษาไทยแบบเบ็ดเสร็จด้วยคอมพิวเตอร์. วิทยานิพนธ์ระดับปริญญาโท จุฬาลงกรณ์มหาวิทยาลัย.
- ปิ่นปิลันธน์ วังชัย. (2553). การท่องเที่ยวโลกปี 2009. จุลสารวิชาการอิเล็กทรอนิกส์ การท่องเที่ยวแห่งประเทศไทย. หน้า 1-11.
- มหาวิทยาลัยสุโขทัยธรรมาราช. (2544). การจัดการสารสนเทศ หน่วยที่ 1-7. นนทบุรี: สำนักพิมพ์มหาวิทยาลัยสุโขทัยธรรมาราช.
- สมาคมนักวิชาการการท่องเที่ยว. (2553). สรุปการสัมมนา การจัดการบริการและการท่องเที่ยวในภาวะวิกฤตระหว่างวันที่ 23-24 พฤศจิกายน 2552 ณ มหาวิทยาลัยสยาม. จุลสารวิชาการอิเล็กทรอนิกส์การท่องเที่ยวแห่งประเทศไทย. หน้า 1-8.
- สิรินดา นุชชัยเหล็ก. (2552). คาดการณ์ผลกระทบจากปัญหาการแพร่ระบาดของโรคไข้หวัดใหญ่สายพันธุ์ใหม่ 2009 (Swine Flu: H1N1). จุลสารวิชาการอิเล็กทรอนิกส์ การท่องเที่ยวแห่งประเทศไทย. หน้า 1-14.
- สุวรรณ บุเหลา, ธรา อังสกุล และจิตินันต์ อังสกุล. (2554). ระบบวางแผนท่องเที่ยวออนไลน์ภายใต้ข้อบังคับด้านความปลอดภัย. งานประชุมวิชาการบัณฑิตศึกษา ครั้งที่ 4. มหาวิทยาลัยเทคโนโลยีสุรนารี นครราชสีมา. กรกฎาคม 7-8. หน้า 142-152.
- George, R. (2003). Tourist's Perceptions of Safety and Security While Visiting Cape Town. **Tourism Management**, 24: 575-585.
- Han, H., Noro, T., and Tokuda, T. (2009). An Automatic Web News Article Contents Extraction System Based on RSS Feeds. **Journal of Web Engineering**, 8 (3): 268-284.
- Howard, R.W. (2009). Risky business? Asking Tourists What Hazards They Actually Encountered in Thailand. **Tourism Management**, 30: 359-365.

-
- Lan, Q. (2010). Extraction of News Content for Text Mining Based on Edit Distance. **Journal of Computational Information Systems**. 6 (11): 3761-3777
- Park, S., Tae, Y. and Park, S. (2006). Self-Organizing n-gram Model for Automatic Word Spacing. **In Proceeding of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL** (pp.633-640). Sydney: Australia.
- Prideaux, B. (2005). Factors Affecting Bilateral Tourism Flows. **Annals of Tourism Research**. 32 (3): 780-801.
- Rittichainuwat, B.N., and Chakraborty, G. (2008). Perceived Travel Risks Regarding Terrorism and Disease: The Case of Thailand. **Tourism Management**. 30: 410-418.
- Zhang, H. Q., Qu, H. and Tang, V.M.Y. (2003). A Case Study of Hong Kong Residents' Outbound Leisure Travel. **Tourism Management**. 25: 267-273.

