

กระบวนการสกัดข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่ และความสามารถในการนำเสนอสารสนเทศด้วยภาพข้อมูลผ่านเว็บไซต์

DATA EXTRACTION PROCESS OF ROAD MAJOR ACCIDENT REPORTS AND USABILITY TESTING OF INFORMATION PRESENTATION BY DATA VISUALIZATION ON WEBSITE

จักรินทร์ สันติรัตนภักดี^{1*}

Chakkarin Santirattanaphakdi^{1}*

สาขาวิชาระบบสารสนเทศคอมพิวเตอร์ คณะบริหารธุรกิจ มหาวิทยาลัยวงษ์ชวลิตกุล
Department of Computer Information System, Faculty of Business Administrator,
Vongchavalitkul University.

*Corresponding author, e-mail: chakkarin_san@vu.ac.th

Received: 9 December 2020; **Revised:** 7 June 2021; **Accepted:** 10 June 2021

บทคัดย่อ

ประเทศไทยมีการจัดเก็บข้อมูลการเกิดอุบัติเหตุ เพื่อนำมาวางแผนความปลอดภัยทางถนน โดยรายงานอุบัติเหตุรายใหญ่ถูกเผยแพร่ในวงจำกัด เนื่องจากต้องสังเคราะห์เนื้อหาด้วยตนเอง ข้อมูลบางส่วนจึงยังมีความผิดพลาด และไม่เป็นมาตรฐานจากผู้ให้ข้อมูลที่แตกต่างกัน ดังนั้นงานวิจัยชิ้นนี้มีวัตถุประสงค์เพื่อ 1) ออกแบบและพัฒนากระบวนการสกัดข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่ รวบรวมข้อมูลบนเว็บไซต์ด้วยเทคนิคการขุดเว็บมาสกัดข้อมูลใน 6 ประเด็น ได้แก่ จังหวัด วันที่ จำนวนผู้เสียชีวิต จำนวนผู้บาดเจ็บ ประเภทและจำนวนของยานพาหนะที่เกิดเหตุ นำเสนอเป็นภาพข้อมูลแก่ผู้ใช้ เพื่อเป็นส่วนหนึ่งในการวางแผนความปลอดภัยทางถนนเชิงพื้นที่หรือตามช่วงเวลา ผลการประเมินความถูกต้องจากกระบวนการสกัดชื่อจังหวัด วันที่ และประเภทของยานพาหนะจากการสร้างคลังคำศัพท์ร่วมกับการรู้จำเอนทิตีที่มีความถูกต้องสูง ขึ้นอยู่กับครอบคลุมของคลังคำศัพท์แต่อย่างไรก็ดี จำนวนผู้เสียชีวิต จำนวนผู้บาดเจ็บ และจำนวนยานพาหนะที่กำหนดรูปแบบอักขระร่วมกับการรู้จำเอนทิตี นับว่ามีความถูกต้องค่อนข้างต่ำ เมื่อเทียบกับการสร้างคลังคำศัพท์ร่วมกับการรู้จำเอนทิตี เนื่องจากรูปแบบการบันทึกข้อมูลที่ไม่เป็นมาตรฐาน 2) ประเมินความสามารถในการนำเสนอสารสนเทศรายงานอุบัติเหตุทางถนนรายใหญ่ด้วยภาพข้อมูลผ่านเว็บไซต์ จากผู้ใช้ 30 คน ด้วยการสุ่มตัวอย่างโดยไม่ใช้ความน่าจะเป็นจากการเลือกกลุ่มตัวอย่างแบบเฉพาะเจาะจง โดยให้ผู้ใช้แต่ละคนทดลองใช้งานแล้วประเมินผลผ่านแบบสอบถามคนละ 3 ครั้งแล้วนำมาหาค่าเฉลี่ย ผลการประเมินภาพรวมอยู่ในระดับดี เมื่อแบ่งผู้ใช้เป็น 3 กลุ่มตามบริบทการใช้งาน ได้แก่ กลุ่มผู้เชี่ยวชาญด้านคอมพิวเตอร์ กลุ่มผู้ทำหน้าที่ข้อมูลจราจร และกลุ่มผู้ใช้ทั่วไป แล้ววิเคราะห์ด้วยสถิติทดสอบเอฟ แบบ LSD พบว่า กลุ่มผู้เชี่ยวชาญด้านคอมพิวเตอร์ และกลุ่มผู้ทำหน้าที่ข้อมูลจราจร มีความแตกต่างด้านการเรียนรู้อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 แสดงถึงประสพการณ์ในการใช้งานระบบด้านการจราจรที่เกี่ยวข้องส่งผลต่อการเรียนรู้มากกว่าประสพการณ์ใช้งานคอมพิวเตอร์และสมาร์ทโฟน ที่แม้จะมีประสพการณ์ใน

การใช้งานคอมพิวเตอร์และสมาร์ทโฟนมากกว่า แต่ผลการประเมินด้านความสามารถในการเรียนรู้ต่ำกว่า อันเป็นแนวทางออกแบบและพัฒนาการสกัดข้อมูลกึ่งมีโครงสร้าง และนำเสนอสารสนเทศด้วยภาพข้อมูลในอนาคต

คำสำคัญ: การสกัดข้อมูล, อุบัติเหตุทางถนน, ความสามารถในการใช้งาน, การนำเสนอภาพข้อมูล

Abstract

Thailand has an accident database to be used for road safety planning. The major accident report is one of the qualitative data waiting to be utilized, because users have to synthesize the content themselves. Therefore published only in a limited. However, some of the information still contains errors and is not standardized from the context of different informants. The research aims 1) to design and develop of a data extraction process. By accessing and collecting data on the website using web scraping techniques for extract data in 6 issues: province, date, number of deaths, number of injured, vehicle type and number of vehicles involved in the accident for presented to user on data visualization and use in the part of road safety planning at the local level in area or periodic. The results of accuracy evaluation from data extraction process in the issue: "province name", "date" and "vehicle type" by corpus-based approach and Named Entity Recognition (NER) method is highly accurate, depend on completely of corpus-based. However, the accuracy of the "number of deaths", "number of injured" and "number of vehicles involved in the accident" by set regular expression for NER method is relatively low compared to corpus-based approach and NER method. Due to a non-standard report format. 2) to usability testing of information presentation by data visualization on website from 30 users determine the sample size on non-probability sampling with purposive sampling, by having each user try it out and evaluate the results through questionnaires 3 times. The usability testing evaluation is good level. When the user was divided into 3 groups according to the context of use: computer user group, traffic information group and general user groups for F-test analysis with LSD found statistically significant difference between the computer experts group and traffic information groups at the 0.05 level in Learnability component. Demonstrate the experience of using relevant traffic systems, which affects the user experience more than the experience of computers and smartphones. At a computer expert group, despite having more using computers and smartphones experience, but the assessment of learnability component is lower than. Which is the design and development of semi-structured data extraction and information presentation by data visualization in the future.

Keywords: Data Extraction, Road Accidents, Usability Testing, Data Visualization

บทนำ

อุบัติเหตุทางถนนนับเป็นหนึ่งในปัญหาสำคัญระดับโลก จากการรายงานขององค์การอนามัยโลก [1] พบว่า มีผู้เสียชีวิตจากการอุบัติเหตุทางถนนเฉลี่ยวันละ 3,700 คน ซึ่งประเทศไทยมีผู้เสียชีวิตจากอุบัติเหตุทางถนนสูงที่สุดเป็นอันดับหนึ่งในเอเชีย และอันดับที่ 9 ของโลก โดยมีผู้เสียชีวิต เฉลี่ย 60 คนต่อวัน จากความรุนแรงของปัญหาดังกล่าว ประเทศไทยเองเห็นชอบให้ศูนย์อำนวยการความปลอดภัยทางถนนจัดทำแผนปฏิบัติการ

ทศวรรษแห่งความปลอดภัยทางถนนกำหนดเป้าหมายคือลดอัตราการเสียชีวิตจากอุบัติเหตุทางถนนต่ำกว่า 10 คนต่อประชากรหนึ่งแสนคนในปีพ.ศ. 2563

ประเทศไทยมีการจัดเก็บข้อมูลการเกิดอุบัติเหตุในช่วงหลายปีที่ผ่านมา และเผยแพร่ให้เข้าถึงอย่างเป็นสาธารณะ เพื่อนำข้อมูลมาช่วยในการวางแผนป้องกันอุบัติเหตุทางถนน แต่การจัดเก็บข้อมูลที่มีหลายหน่วยงานรับผิดชอบ และมีวัตถุประสงค์ในการเก็บข้อมูลแตกต่างกัน [2] ส่งผลให้ข้อมูลมีคลาดเคลื่อน ทำให้ข้อมูลอุบัติเหตุของประเทศไทยขาดความน่าเชื่อถือ จึงเป็นที่มาของการบูรณาการข้อมูลจากกระทรวงสาธารณสุข สำนักงานตำรวจแห่งชาติ และบริษัทกลางคุ้มครองผู้ประสบภัยจากรถ จำกัด ดังนั้นข้อมูลที่นำเชื่อถือและยังรอการนำไปใช้ประโยชน์อีกรูปแบบหนึ่ง นั่นคือการรายงานอุบัติเหตุรายใหญ่ (Major Accident) [3] ที่มีผู้เสียชีวิต 2 รายขึ้นไป หรือผู้บาดเจ็บ 4 คนขึ้นไป หรือผู้บาดเจ็บและผู้เสียชีวิตรวม 4 คน ขึ้นไป ตลอดจนอุบัติเหตุทางถนนที่เป็นที่สนใจของประชาชน เช่น บุคคลสำคัญหรือบุคคลสาธารณะเกิดบาดเจ็บ พิการขาตัวที่โรงพยาบาล หรือเสียชีวิต หรือรถน้ำมัน รถสารเคมีเกิดอุบัติเหตุ จนเกิดเพลิงไหม้ หรือสารเคมีฟุ้งกระจายไปสู่ชุมชน เป็นต้น ที่รายงานมายังศูนย์รับแจ้งอุบัติเหตุและประสานการช่วยเหลือผู้ประสบภัยจากรถ บริษัทกลางคุ้มครองผู้ประสบภัยจากรถ จำกัด เผยแพร่ต่อสาธารณะผ่านเว็บไซต์ www.thairsc.com เป็นข้อมูลเพื่อนำไปเยียวยาผู้ประสบภัยจากรถให้ได้รับการคุ้มครองตามกฎหมาย เพื่อเป็นแนวทางในการป้องกันและลดการเกิดอุบัติเหตุทางถนน ในรูปแบบของข้อมูลประเภทกึ่งมีโครงสร้าง (Semi-Structure Data) จากการกำหนดหัวข้อในการบันทึกข้อมูล โดยแต่ละช่องเปิดโอกาสให้ผู้รายงานสามารถใส่ข้อมูลตามสถานการณ์ที่เกิดขึ้นจริง ทำให้ในแต่ละครั้งข้อมูลในการรายงานขาดความถูกต้อง ไม่เป็นมาตรฐาน และแตกต่างกันไปตามบริบทของผู้รายงาน จึงยากต่อการวิเคราะห์และนำไปใช้ประโยชน์ ผู้วิจัยทราบถึงปัญหา และเล็งเห็นถึงความสำคัญของข้อมูล จึงออกแบบและพัฒนากระบวนการสกัดข้อมูลรายงานอุบัติเหตุรายใหญ่จากเว็บไซต์ เพื่อสกัดข้อมูลที่เพิ่มขึ้นอย่างมากในปัจจุบัน และมีแนวโน้มที่จะเพิ่มขึ้นเรื่อย ๆ ในอนาคต ด้วยวิธีการจัดประเภท การหาความหมายหรือใจความสำคัญจากเอกสารแบบกึ่งมีโครงสร้าง หรือไม่มีโครงสร้าง เรียกว่าการรู้จำชื่อเอนทิตี (Named Entity Recognition : NER) [4] เป็นหนึ่งในวิธีการสกัดข้อมูลที่ได้รับความนิยมอย่างมากในปัจจุบัน โดยใช้วิธีการทำป้ายลำดับ (Sequence Labeling) ในการกำหนดชื่อเฉพาะ ในลักษณะของการหาและระบุตำแหน่งของข้อความ ตลอดจนแยกประเภทและส่วนประกอบของข้อความให้อยู่ในประเภทที่ต้องการ รูปแบบที่ใช้นำแบ่งเป็น 2 วิธี คือ 1) แบบใช้กฎ (Rule-Based) ข้อดีคือมีความถูกต้องสูงเพราะมีผู้เชี่ยวชาญในการระบุหน้าที่ชัดเจน แต่มีข้อเสียกับข้อมูลที่ต้องการความเชี่ยวชาญเฉพาะทางจะหาได้ยาก นอกจากนี้ยังใช้เวลาค่อนข้างนานตามความซับซ้อนของโมเดล ประกอบด้วยกำหนดยุทธวิธีแบบอักขระ (Regular Expression) การกำหนดด้วยพจนานุกรม (Dictionary-Based) และการกำหนดโดยไวยากรณ์ (Linguistic Grammar) และ 2) แบบการเรียนรู้ของเครื่องคอมพิวเตอร์ (Machine Learning) โดยการสร้างตัวแปลงเอกสารอัตโนมัติ (Wrapper) เพื่อแปลงเอกสารเอชทีเอ็มแอลอยู่ในรูปแบบที่ผู้ใช้กำหนดไว้ [5]

ปัจจุบันมีงานวิจัยที่นำมุกสกัดข้อมูลด้วยเทคนิคการรู้จำชื่อเอนทิตีเป็นจำนวนมาก ซึ่งถูกประยุกต์ใช้อย่างกว้างขวางในศาสตร์หลายแขนง หลายภาษา แต่ละงานก็มีเทคนิคและวิธีการแตกต่างกันออกไป เช่น นำการรู้จำชื่อเอนทิตีในการจัดการกับปัญหาของคำภาษาจีนที่ไม่รู้จัก [6] โดยแบ่งเป็นประโยคตามลำดับของคำที่รู้จักแล้วใช้คลังคำศัพท์ HMM-based ในการกำหนดให้แต่ละคำเป็นคำที่มีความหมายด้วยแท็กไฮบริดที่เหมาะสมในการสกัดโครงสร้างข้อมูลประเภทนิติบุคคล และหมวดหมู่ของนิติบุคคล เป็นผลให้ระบบสามารถปรับปรุงการเพิ่มคำ ชุดวลีหรือรูปแบบคำลงในคลังพจนานุกรมภาษาได้ โดยไม่ลดประสิทธิภาพของชุดข้อมูลฝึกสอน (Training Data) และการติดแท็ก ภายหลัง [7] มีการนำเสนอวิธีการแยกเอนทิตีชื่อภาษาไทย เช่น ชื่อบุคคล ชื่อองค์กร สถานที่ วันที่ และเวลา ตลอดจนวลี จากข้อความโดยไม่ต้องใช้การแบ่งส่วนคำและการติดแท็กบางส่วน พบว่า การใช้พจนานุกรมภาษาไทยขนาดใหญ่และชุดเทมเพลตการจับคู่รูปแบบที่กำหนดไว้ล่วงหน้าสามารถตรวจหาเอนทิตีที่มี

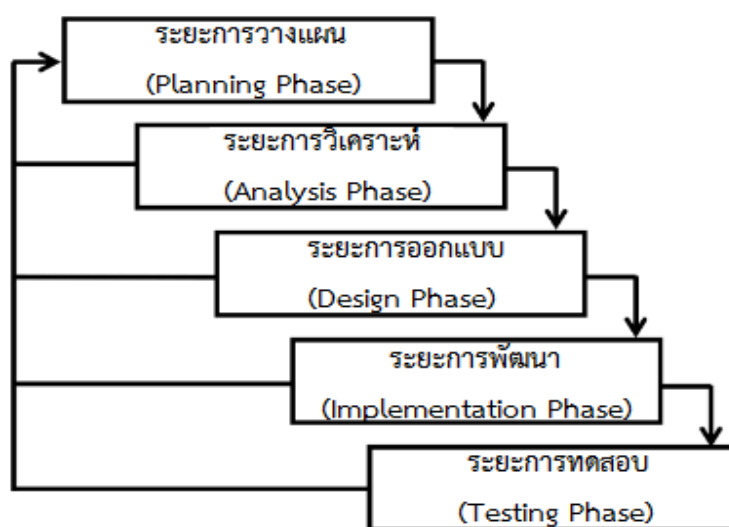
ชื่อถูกต้องประมาณ 68-100% ต่อมา [8] นำการรู้จำชื่อเอนทิตีมาสกัดข้อมูลของข้อความยาวไม่เกิน 280 ตัวอักษรหรือทวีต (Tweet) ที่มักเกิดปัญหาเกี่ยวกับข้อมูลที่ทวีตไม่เพียงพอ และการขาดชุดข้อมูลฝึกสอน ด้วยองค์ประกอบหลักสามประการ คือ 1) กระบวนการเตรียมข้อมูลก่อนนำไปประมวลผล โดยแก้ไขคำที่มีรูปแบบไม่ถูกต้องโดยใช้แบบจำลองเชิงเส้น 2) ใช้วิธีการเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors: KNN) ในการจัดแบ่งกลุ่มป้ายลำดับจากเอกสารต่าง ๆ และทำป้ายลำดับโดยใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์ (Conditional Random Fields: CRF) เพื่อค้นหาและระบุข้อมูลที่ต้องการ และ 3) ใช้กรอบการเรียนรู้แบบกึ่งมีผู้สอน (Semisupervised Learning) ที่ช่วยลดการใช้ชุดข้อมูลฝึกสอน ผลลัพธ์แสดงให้เห็นถึงประสิทธิภาพของสกัดข้อมูลที่เพิ่มมากขึ้น จากนั้นมีการใช้การรู้จำชื่อเอนทิตีในการสกัดภาษาอูรดู (Urdu) [9] เป็นภาษาหนึ่งในภาษากลุ่มอินโด-อารยัน เริ่มจากพัฒนาคลังคำศัพท์ภาษาอูรดู พร้อมใส่คำอธิบายประกอบ จากนั้นติดแท็กอย่างน้อย 2 แท็กเพื่อจำแนกประเภท ร่วมกับการฝังคำใหม่ 6 คำโดยใช้เทคนิคแบบ fastText, Word2vec และ Glove บนข้อความภาษาอูรดู โดยใช้กระบวนการเรียนรู้เชิงลึกแบบโครงข่ายประสาทเทียม (Neural Network) และโครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network) สำหรับการรู้จำชื่อเอนทิตี ผลการทดลอง พบว่า เทคนิคการเรียนรู้เชิงลึกมีผลกระทบของการฝังคำและรูปแบบต่าง ๆ ของชุดข้อมูล นอกจากนี้ยังมีการพัฒนาระบบในการดึงและเลือกข้อมูลบุคคลที่มีชื่อเสียงของประเทศไทยจากเว็บไซต์แบบอัตโนมัติ [10] เนื่องจากข้อมูลแบบดั้งเดิมจะถูกรวบรวมและคัดเลือกด้วยมือ ซึ่งค่อนข้างใช้เวลานานและมักไม่สามารถปรับปรุงข้อมูลได้ เนื่องจากกลุ่มบุคคลดังกล่าวในประเทศไทยมีจำนวนมากตลอดจนพบความคลุมเครือและความขัดแย้งที่อาจเกิดขึ้นระหว่างแหล่งข้อมูลบนอินเทอร์เน็ต งานวิจัยนี้ใช้การจับคู่รูปแบบและกฎการเชื่อมโยง เพื่อแยกวันเดือนปีเกิด ส่วนสูงและน้ำหนักของคนดังจากเว็บไซต์ ผลลัพธ์พบว่าระบบสามารถสกัดข้อมูลบุคคลที่มีชื่อเสียงได้มากกว่าหลายเว็บไซต์ในประเทศไทย และการประยุกต์ใช้เทคนิคการเรียนรู้จำเอนทิตี เพื่อสกัดชื่อกษัตริย์ผู้สร้างวัดในสมัยล้านนาจากเอกสารประวัติวัดในรูปแบบไฟล์ข้อความ [11] เริ่มจากการรวบรวมเอกสารข้อมูลประวัติวัดในเขตอำเภอเมืองเชียงใหม่ โดยนำเข้าข้อมูลไฟล์ข้อความ เพื่อตัดคำภาษาไทยที่เน้นการแยกคำแต่ละคำออกจากกันโดยคงความหมายเดิมด้วยโปรแกรม LexTo ด้วยการกำหนดชื่อกษัตริย์และคำสำคัญที่บ่งบอกถึงการสร้างวัดไว้ในพจนานุกรมข้อมูล เพื่อใช้ในการอ้างอิง เช่นเดียวกับ [12] ในโดเมนร้านอาหาร โรงแรม การเลือกซื้อสินค้า และชุดข้อมูลการท่องเที่ยวของประเทศไทยที่รวบรวมจากเว็บไซต์ด้วยการรู้จำชื่อเอนทิตี การจัดประเภทข้อความ และการสรุปข้อความ เริ่มจากการรวบรวมข้อมูลการท่องเที่ยวและสร้างคลังคำศัพท์ ได้แก่ การแยกประโยคความสัมพันธ์ และการแยกเอนทิตีชื่อในการติดแท็กสำหรับการสร้างชุดข้อมูลฝึกสอนที่เหมาะสม จากนั้นจะสร้างแบบจำลองการรับรู้ของประเภทเอนทิตี เพื่อแยกชื่อสถานที่ สิ่งอำนวยความสะดวก ตลอดจนประเภทความสัมพันธ์ และจัดประเภทบทวิจารณ์หรือสรุปบทวิจารณ์ โดยใช้ SpaCy และ BERT เป็นเครื่องมือเปรียบเทียบประสิทธิภาพของผลลัพธ์ จากการทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้องพบว่า มีหลากหลายแนวคิดในการพัฒนาวิธีการเพิ่มประสิทธิภาพในการสกัดข้อมูล งานวิจัยชิ้นนี้เป็นการออกแบบและพัฒนากระบวนการสกัดข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่จากเว็บไซต์ ด้วยการสร้างคลังข้อมูลคำศัพท์และกำหนดรูปแบบอักขระสำหรับสกัดข้อมูลด้วยการรู้จำชื่อเอนทิตี แล้วนำผลลัพธ์ที่ได้มานำเสนอสารสนเทศด้วยภาพข้อมูลผ่านเว็บไซต์ ที่สามารถนำมาเป็นส่วนหนึ่งในการวางแผนป้องกันและลดอุบัติเหตุในลักษณะนโยบายเชิงพื้นที่ หรือกำหนดนโยบายตามช่วงระยะเวลา ตลอดจนเสนอแนวทางการออกแบบและพัฒนาการสกัดข้อมูลประเภทกึ่งมีโครงสร้าง และนำเสนอสารสนเทศในอนาคต

วัตถุประสงค์ของการวิจัย

1. เพื่อออกแบบและพัฒนากระบวนการสกัดข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่
2. เพื่อประเมินความสามารถในการนำเสนอสารสนเทศรายงานอุบัติเหตุทางถนนรายใหญ่ด้วยภาพข้อมูลผ่านเว็บไซต์

วิธีดำเนินการวิจัย

งานวิจัยชิ้นนี้ใช้วงจรการพัฒนาระบบรูปแบบน้ำตกแบบย้อนกลับได้ (Adapted Waterfall Model) [13] ที่ปรับปรุงมาจากรูปแบบน้ำตกแบบดั้งเดิม [14] โดยรวมคุณสมบัติแบบทวนซ้ำเป็นรอบเข้าไป เพื่อตรวจสอบว่าแต่ละส่วนสามารถทำงานร่วมกันได้อย่างไม่มีปัญหาเป็นกรอบในการดำเนินงาน ประกอบด้วย 5 ระยะ ดังภาพที่ 1



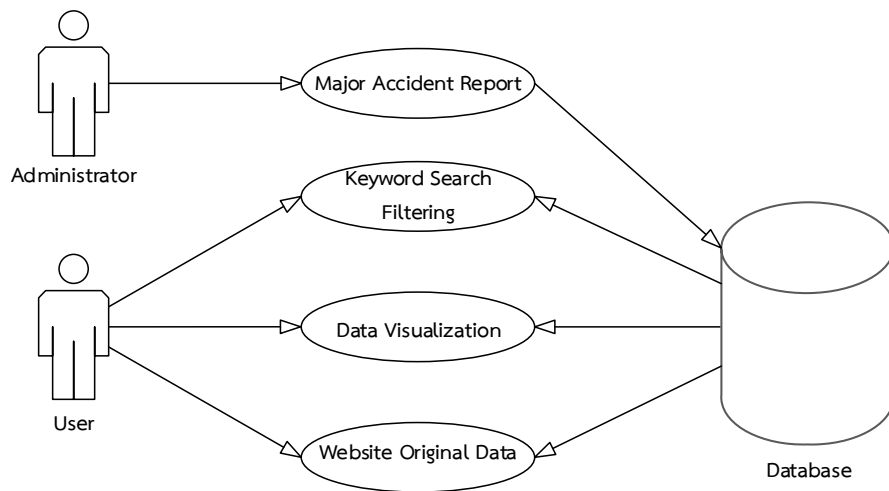
ภาพที่ 1 วงจรการพัฒนาระบบรูปแบบน้ำตกแบบย้อนกลับ

1. ระยะการวางแผน รวบรวมข้อมูลจากการศึกษาระบบจัดเก็บข้อมูลการเกิดอุบัติเหตุของประเทศไทยที่มีการเผยแพร่ผ่านเว็บไซต์ คือ 1) ระบบรายงานสถานการณ์อุบัติเหตุบนโครงข่ายคมนาคมของกระทรวงคมนาคม (Thailand Road Accident Management System: TRAMS) ที่แสดงตำแหน่งและรายละเอียดของการเกิดอุบัติเหตุบนระบบสารสนเทศภูมิศาสตร์ และแสดงข้อมูลสถิติต่าง ๆ และ 2) ระบบบันทึกข้อมูลของบริษัทกลางคุ้มครองผู้ประสบภัยจากรถ จำกัด ที่สามารถกำหนดจุดเกิดเหตุผู้เสียชีวิต เพื่อหาสถิติการเกิดอุบัติเหตุในแต่ละจุดจากข้อมูลผู้ใช้สิทธิตามพระราชบัญญัติคุ้มครองผู้ประสบภัยจากรถของบริษัทกลางคุ้มครองผู้ประสบภัยจากรถ จำกัด ทุกจังหวัดทั่วประเทศ ร่วมกับการสังเคราะห์เอกสาร และงานวิจัยที่เกี่ยวข้อง

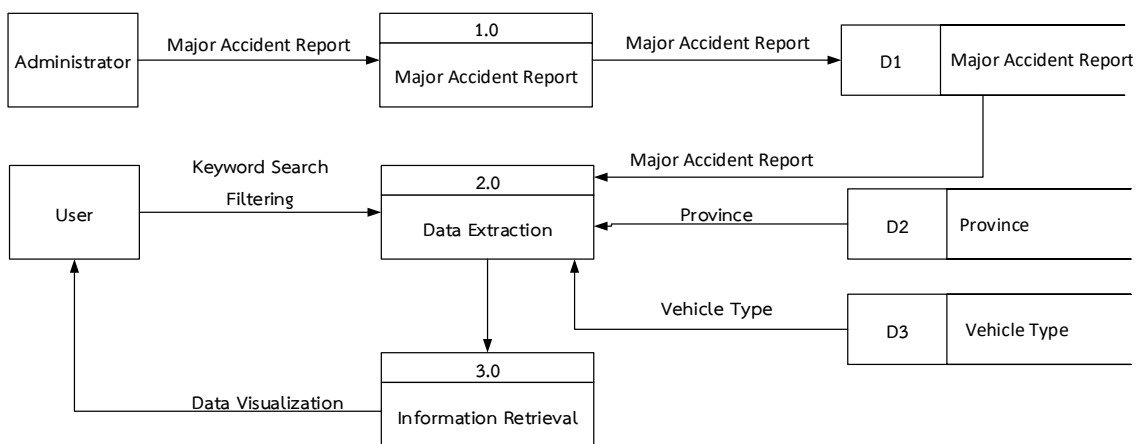
2. ระยะการวิเคราะห์ จากการศึกษาและจัดเก็บข้อมูลการเกิดอุบัติเหตุร่วมกับการสังเคราะห์เอกสาร และงานวิจัยที่เกี่ยวข้อง พบว่า ส่วนใหญ่เน้นจัดเก็บเพียงข้อมูลเชิงปริมาณเท่านั้น แต่รายงานอุบัติเหตุขนาดใหญ่บนเว็บไซต์ที่เป็นข้อมูลเชิงคุณภาพนั้นเพียงแค่เผยแพร่แก่ผู้ใช้ในวงจำกัด และมีได้ผ่านการประมวลผลเพื่อนำไปใช้ประโยชน์ จึงนำทฤษฎีการวิเคราะห์และออกแบบระบบเชิงวัตถุ โดยใช้แบบจำลองกรณีใช้งานระบบ (Use Case Diagram) [13] ที่แสดงมุมมองภาพรวมในพฤติกรรมของระบบเบื้องต้น และแสดงแผนภาพความต้องการทางด้านซอฟต์แวร์ที่เป็นฟังก์ชันการทำงานของระบบในรูปแบบสัญลักษณ์ของยูสเคส ดังภาพที่ 2 ร่วมกับการ

ออกแบบแผนภาพกระแสข้อมูล (Data Flow Diagram : DFD) [14] เพื่อแสดงความสัมพันธ์ระหว่างข้อมูลกับกระบวนการทำงานและคำอธิบายกระบวนการทำงาน (Process Description) ดังภาพที่ 3

ข้อมูลรายงานอุบัติเหตุรายใหญ่ที่รวบรวมจากเว็บไซต์ www.thairsc.com จะนำข้อมูลเข้าสู่กระบวนการสกัดข้อมูลโดยผู้วิจัยตามกระบวนการที่ได้ออกแบบและพัฒนาไว้ด้วยการสร้างคลังคำศัพท์ร่วมกับการรู้จำเอนทิตีใน 6 ประเด็น ได้แก่ จังหวัด วันที่ จำนวนผู้เสียชีวิต จำนวนผู้บาดเจ็บ ประเภทและจำนวนของยานพาหนะที่เกิดเหตุ เพื่อนำเสนอสารสนเทศด้วยภาพข้อมูลผ่านเว็บไซต์ แบ่งเป็นการค้นหาข้อมูลเบื้องต้น (Basic Search) ผู้ใช้จะระบุเงื่อนไขวันเริ่มต้น วันสิ้นสุด และเลือกจังหวัดที่ต้องการแสดงผลในแต่ละครั้ง หรือหากผู้ใช้ต้องการผลลัพธ์ที่มีความละเอียดมากขึ้น (Advanced Search) ก็สามารถระบุเงื่อนไขชนิดและจำนวนยานพาหนะในที่เกิดเหตุ ตลอดจนจำนวนผู้บาดเจ็บและจำนวนผู้เสียชีวิตร่วมด้วยได้ จากนั้นจะทำการสกัดข้อมูลตามกระบวนการที่ได้ออกแบบและพัฒนาไว้ และแสดงผลแก่ผู้ใช้ในลักษณะการนำเสนอภาพข้อมูลแบบมีปฏิสัมพันธ์กับผู้ใช้ ตลอดจนผู้ใช้สามารถเข้าถึงเนื้อหาต้นฉบับผ่านทางลิงก์รายงานแต่ละรายการ



ภาพที่ 2 แบบจำลองกรณีปฏิบัติงานกระบวนการสกัดข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่ และนำเสนอสารสนเทศด้วยภาพข้อมูลผ่านเว็บไซต์



ภาพที่ 3 แผนภาพกระแสข้อมูลกระบวนการสกัดข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่ และความสามารถในการนำเสนอสารสนเทศด้วยภาพข้อมูลผ่านเว็บไซต์

3. ระยะเวลาออกแบบ กำหนดกระบวนการสกัดข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่ ในลักษณะเว็บแอปพลิเคชันที่ปฏิสัมพันธ์กับผู้ใช้ผ่านเว็บเบราว์เซอร์ ด้วยภาษา PHP ร่วมกับ Javascript และ AJAX สำหรับพัฒนาเว็บเพจแบบไดนามิกสีในการสกัดข้อมูลรายงานอุบัติเหตุขนาดใหญ่ โดยใช้ MySQL เป็นฐานข้อมูล เพื่อนำข้อมูลดังกล่าวมาสร้างการนำเสนอสารสนเทศด้วยภาพข้อมูลผ่านเว็บไซต์ด้วย Google Chart และผู้วิจัยคำนึงถึงการใช้งานที่หลากหลาย จึงออกแบบให้เว็บไซต์รองรับการทำงานบนขนาดหน้าจอที่แตกต่างกัน เช่น คอมพิวเตอร์ สมาร์ทโฟน และแท็บเล็ตในเว็บไซต์เดียวตามหลักการออกแบบเรสปอนซีฟเว็บไซต์ (Responsive Web Design) แบบ Responsive Retrofitting [15] หมายถึง การแปลงเว็บไซต์ที่ทำขึ้นมาเพื่อรองรับหน้าจอคอมพิวเตอร์ด้วยการเพิ่ม CSS3 Media Query เพื่อให้รองรับการแสดงผลหน้าจอบนอุปกรณ์อื่น ที่มีจุดเด่นอยู่ที่ความเร็วในการพัฒนา เนื่องจากไม่ต้องใช้เวลาในการออกแบบใหม่สำหรับอุปกรณ์ชนิดอื่น จึงได้รับความนิยมอย่างสูงในปัจจุบัน

4. ระยะเวลาพัฒนา แบ่งเป็น 5 ขั้นตอน ตามลำดับการทำงาน มีรายละเอียดดังนี้

4.1 การเข้าถึงข้อมูลรายงานอุบัติเหตุรายใหญ่เพื่อนำมาประมวลผล เนื่องจากมีข้อมูลในเว็บไซต์เป็นจำนวนมาก การจะเข้าถึงข้อมูลที่ละรายการด้วยตนเองไม่ใช่เรื่องง่าย อีกทั้งรูปแบบข้อมูลที่อยู่บนเว็บไซต์นั้นถูกออกแบบมาเพื่อแสดงผลให้กับผู้ใช้เท่านั้น ไม่สามารถนำมาวิเคราะห์ได้โดยตรง จึงนำเทคนิคการขูดเว็บ (Web Scraping) [16] ที่เสมือนกับว่าเข้าใช้เว็บไซต์ด้วยตนเอง แต่ในทางกลับกันจะเขียนคำสั่ง หรือใช้โปรแกรมสำเร็จรูปในการเข้าถึงเว็บไซต์ และคัดลอกข้อมูลในตำแหน่งที่ระบุมาจัดเก็บในรูปแบบของไฟล์ที่รอกการนำไปประมวลผล อย่างไรก็ตามการต้องระงับมิให้สร้างภาระในการทำงานของเครื่องแม่ข่ายจนอาจก่อให้เกิดความเสียหาย และการอนุญาตให้เข้าถึงและคัดลอกข้อมูลบนเว็บไซต์ ดังนั้นผู้วิจัยทำการทดสอบว่าเว็บไซต์ www.thairsc.com นั้นอนุญาตการใช้งานของเว็บครอว์เลอร์ (Web Crawler) [17] ซึ่งเป็นโปรแกรมอัตโนมัติที่ใช้ในการเก็บข้อมูลจากเว็บไซต์ด้วยการเขียนโปรแกรมภาษา R ด้วยไลบรารี robots.txt ปรากฏผลลัพธ์คือการคืนค่า TRUE แสดงถึงเว็บไซต์ดังกล่าวสามารถใช้เว็บครอว์เลอร์ในการเก็บข้อมูลได้

งานวิจัยชิ้นนี้เข้าถึงข้อมูลรายงานอุบัติเหตุรายใหญ่บนเว็บไซต์ www.thairsc.com ด้วยการเขียนโปรแกรมภาษา R [18] ผ่านไลบรารี Rvest และ xml2 เพื่อเข้าถึงเว็บเพจแต่ละหน้าตามที่กำหนดไว้แบบวนลูป จากนั้นจะคัดลอกข้อมูลที่อยู่รูปแบบรายงานภาษาอังกฤษที่ปรากฏบนเว็บไซต์ตั้งแต่ 1 มกราคม 2559 – 31 มีนาคม 2563 รวมทั้งสิ้น 3,025 รายการ แล้วนำมาบันทึกเป็นไฟล์ CSV (Comma Separated Value) ประกอบด้วยลำดับ ส่วนหัวของรายงาน ส่วนรายละเอียดของรายงาน และลิงก์การเข้าถึง ดังภาพที่ 4 เพื่อนำไปประมวลผลในขั้นตอนต่อไป

48023	accident: 2 cars collided, 3 deaths, 5 injuries, si kiew district, nakhon ratchasima (15-03-2020, 04.50 a.m.)	accident: 2 cars collided, 3 deaths, 5 injuries, si kiew district, nakhon ratchasima (15-03-2020, 04.50 a.m.) accident notification center/road accident victims protection co., ltd. department of disaster prevention and mitigation accident report/notification no: e-accident : 63/001/0004402 source: js100. date of claim: sunday march 15th, 2020, at about 08.00 a.m. date of accident: sunday march 15th, 2020, at about 04.50 a.m. accident point: highway 2, mitrphap road, km.104-105, u-turn bridge are, mitrphap, si kiew district, nakhon ratchasima. accident condition: straightway weather condition: rainy. responsibility area: si kiew police station tel. 0-4441-1010 enquiry officer/case owners: - tel. -accident scene: accident: 2 cars collided. photo by: - motor vehicle(s) involved in the accident: 1. sedan car, honda, civic, grey color, license plate registration no. xxx-xxx maha sarakhm. 2. pick-up car, mitsubishi, white color, license plate registration no. xxx-xxxx udon thani. list of deaths: 3 deaths (male: 2, female: 1) list of injured: 5 injuries (male: 3, female: 2) taken to si kiew hospital. injuries identification: -cause of the accident: (the real causes of the accident is being investigated.) insurance assistance and coordination: 1. license plate registration no. xxx-xxx maha sarakhm, compulsory insurance is not found! 2. license plate registration no. xxx-xxxx udon thani, compulsory insurance is not found! hotline(s): si kiew police station (tel. 0-4441-1010) promtham rescue, si kiew district (tel. 0-4498-4115) si kiew hospital (tel. 0-4498-6240) department of disaster prevention and mitigation (call center 1784) road accident victims protection company limited, ioc and the office of insurance commission (oic)	http://www.thairsc.com/th/bigaccdetail.aspx?l=en&qid=48023
-------	---	---	--

ภาพที่ 4 รายงานอุบัติเหตุทางถนนรายใหญ่จาก www.thairsc.com ที่อยู่ในรูปของ CSV

4.2 การเปลี่ยนรูปแบบข้อมูลก่อนนำไปประมวลผล ในเบื้องต้นผู้วิจัยทำการคัดกรองข้อมูล (Data Filtering) เพื่อให้ได้ข้อมูลที่มีความสมบูรณ์ โดยตรวจสอบโครงสร้างของข้อมูลจากขั้นตอนที่ผ่านมา พบว่า รายงานอุบัติเหตุรายใหญ่ที่อยู่ในรูปแบบภาษาอังกฤษจำนวนหนึ่งยังขาดความสมบูรณ์ในด้านเนื้อหาเมื่อเทียบกับรายงานในรูปแบบภาษาไทย เช่น รายงานอุบัติเหตुरูปแบบภาษาอังกฤษแต่รายงานด้วยภาษาไทย หรือรายงานด้วยรูปภาพอุบัติเหตุแต่ไม่มีรายละเอียดอุบัติเหตุ ตลอดจนมีข้อมูลเพียงส่วนหัวเท่านั้นแต่ไม่ปรากฏในส่วนรายละเอียด เป็นต้น ดังนั้นจึงพิจารณาตัดข้อมูลที่ตรงตามเงื่อนไขออก คงเหลือข้อมูลทั้งสิ้น 1,353 รายการ จากนั้นนำข้อมูลเข้าสู่โปรแกรม Rstudio อีกครั้ง เพื่อเข้าสู่กระบวนการเตรียมข้อมูลรูปแบบข้อความก่อนนำไปประมวลผลด้วยการเขียนโปรแกรมภาษา R เพื่อทำความสะอาดข้อความ (Text Cleaning) โดยการจัดการกับเครื่องหมาย หรืออักขระพิเศษที่ใช้ในการแสดงผล ค้นหาและแทนที่ตัวอักษรภาษาไทยให้เป็นภาษาอังกฤษทั้งหมด ปรับรูปแบบของวันที่ให้อยู่ในรูปแบบมาตรฐาน ตลอดจนตัดข้อความส่วนท้ายของทุก ๆ รายการออก เนื่องจากเป็นเพียงรายละเอียดที่ตั้งและช่องทางการติดต่อเพื่อให้เหลือเฉพาะข้อความที่พร้อมนำไปประมวลผล

4.3 การสร้างคลังคำศัพท์ เมื่อพิจารณารายละเอียดข้อมูลรายงานอุบัติเหตุรายใหญ่ พบว่า มีจำนวนมากที่ยังมีความผิดพลาด เช่น ชื่อจังหวัด จาก “Khon Kaen” เป็น “Khon Kean” รวมถึงชนิดของยานพาหนะในที่เกิดเหตุที่เรียกแตกต่างกันไปตามบริบทของผู้รายงาน ทั้ง “Bike” และ “Bicycle” ซึ่งเป็นอุปสรรคต่อการนำไปใช้ประโยชน์ ดังนั้นจึงนำข้อมูลดังกล่าวมาสร้างคลังข้อมูลคำศัพท์รายชื่อจังหวัดในประเทศไทย อันจะนำมาเป็นเงื่อนไขในการค้นคืนสารสนเทศ จากนั้นนำประเภทของยานพาหนะตามการจำแนกประเภทจากรายงานการวิเคราะห์สถานการณ์อุบัติเหตุทางถนนของกระทรวงคมนาคม 10 ลำดับแรก [19] มาสร้างคลังคำศัพท์ และกำหนดน้ำหนักของคำศัพท์ตามจำนวนตัวอักษร เพื่อเป็นเงื่อนไขการเรียงลำดับในการประมวลผลจากจำนวนตัวอักษร

4.4 การสกัดข้อมูล ใน 6 ประเด็น ได้แก่ จังหวัด วันที่ จำนวนผู้เสียชีวิต จำนวนผู้บาดเจ็บ ประเภท และจำนวนของยานพาหนะที่เกิดเหตุ โดยจะไม่นำข้อมูลส่วนบุคคลของผู้ประสบเหตุมาเป็นส่วนหนึ่งของผลลัพธ์จากการสกัดข้อมูล โดยสร้างเงื่อนไขในการค้นหารายชื่อจังหวัดจากส่วนหัวของรายงานอุบัติเหตุแต่ละรายการ หากไม่พบข้อมูลจะไปค้นหาในส่วนรายละเอียด โดยยึดตามความถี่ที่พบมากที่สุด ในเขตข้อมูล และสร้างเงื่อนไขในการค้นหาประเภทของยานพาหนะในที่เกิดเหตุจากส่วนรายละเอียดแบบวนลูป เริ่มจากคำศัพท์ที่มีความยาวมากที่สุดไปจนถึงน้อยที่สุด ด้วยการเขียนโปรแกรมภาษา PHP เพื่อค้นหาและระบุตำแหน่งของข้อความตามหลักการรู้จำชื่อเอนทิตี โดยทำงานตามลำดับดังนี้

4.4.1 สกัดข้อมูลรายชื่อจังหวัดจากส่วนหัวของรายงาน ตามที่ระบุไว้ในคลังข้อมูลคำศัพท์ที่สร้างขึ้นแบบวนลูป หากมีผลลัพธ์มากกว่า 1 ค่าจะยึดตามความถี่ที่พบมากที่สุด ในเขตข้อมูล

4.4.2 สกัดข้อมูลวันที่จากส่วนหัวของรายงาน ตามรูปแบบวันที่-เดือน-ปี หากไม่ปรากฏผลลัพธ์จะทำการค้นหาต่อไปในส่วนรายละเอียด โดยยึดข้อมูลหลังแท็ก Date of Accident: จนถึงช่องว่าง แล้วทำการแปลงวันที่ให้อยู่ในรูปแบบปี-เดือน-วัน ก่อนนำไปแสดงผล

4.4.3 สกัดข้อมูลจำนวนผู้เสียชีวิต จากส่วนรายละเอียด โดยยึดตัวเลขหลังแท็ก List of Deaths: จนถึงช่องว่าง เช่นเดียวกับจำนวนผู้บาดเจ็บที่ยึดตัวเลขหลังแท็ก List of Injured: หากมีตัวอักษรอยู่รวมในข้อมูล จะทำการกรองเฉพาะตัวเลขเท่านั้น

4.4.4 สกัดข้อมูลประเภทของยานพาหนะในที่เกิดเหตุจากส่วนรายละเอียด โดยค้นหาข้อมูลทั้งหมดหลังแท็ก Motor Vehicle(s) involved in the accident จนถึงแท็ก List of Injured: กำหนดเป็นชุดข้อมูลสำหรับค้นหาประเภทของยานพาหนะตามที่ระบุไว้ในคลังข้อมูลคำศัพท์ที่สร้างขึ้นแบบวนลูป ผลลัพธ์จะมีจำนวนไม่แน่นอน กำหนดขั้นต่ำต้องมีข้อมูลอย่างน้อย 1 ค่า ขึ้นอยู่กับจำนวนยานพาหนะในที่เกิดเหตุ หากมีผลลัพธ์

มากกว่า 1 ค่าจะเก็บข้อมูลแบบอาร์เรย์ เพื่อนำประเภทของยานพาหนะมาทำการวนลู่อีกครั้ง เพื่อหาจำนวนยานพาหนะแต่ละชนิดที่ปรากฏในชุดข้อมูล หากมีตัวอักษรอยู่รวมในข้อมูล จะทำการกรองเฉพาะตัวเลขเท่านั้น

4.5 การนำเสนอข้อมูลในลักษณะรูปภาพ เพื่อถ่ายทอดข้อมูลเชิงปริมาณให้ง่ายต่อการทำความเข้าใจ งานวิจัยชิ้นนี้ใช้เทคนิคการนำเสนอภาพข้อมูลด้วย Google Charts [20] ที่ทำงานผ่านส่วนต่อประสานโปรแกรมประยุกต์ (Application Program Interface : API) โดยจะส่งข้อมูลไปประมวลผลที่เครื่องแม่ข่าย และรับข้อมูลมาแสดงผลบนเว็บไซต์ เพื่อนำเสนอภาพข้อมูลแบบจัดกลุ่ม (Classification) ด้วยแผนภูมิวงกลม (Pie Chart) ในการแสดงผลสถิติยานพาหนะที่เกิดอุบัติเหตุ และแผนที่ประเทศไทย (Geographical) ในการแสดงผลเปรียบเทียบสถิติการเกิดอุบัติเหตุ จำนวนผู้บาดเจ็บ และจำนวนผู้เสียชีวิตแต่ละจังหวัด [21] การแสดงผลจะใช้หลักสีตามลำดับ (Sequential Color) [22] ในงานวิจัยชิ้นนี้เลือกใช้สีขาว สีเหลือง สีส้ม และสีแดงในการแสดงผล โดยไล่ลำดับความเข้มของสีจากน้อยที่สุดไปยังมากที่สุดตามปริมาณการเกิดอุบัติเหตุ

5. ระยะเวลาทดสอบ งานวิจัยนี้เป็นการวิจัยเชิงปริมาณ (Quantitative Research) โดยใช้วิธีวิจัยรูปแบบการสำรวจ (Survey Research) เพื่อประเมินความสามารถของกระบวนการสกัดข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่ และความสามารถในการนำเสนอสารสนเทศด้วยภาพข้อมูลผ่านเว็บไซต์ โดยผู้วิจัยกำหนดขั้นตอน ดังนี้

5.1 การประเมินความถูกต้องของกระบวนการสกัดข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่ โดยทำการวัดความถูกต้องของผลลัพธ์จากกระบวนการสกัดข้อมูลใน 6 ประเด็น ได้แก่ จังหวัด วันที่ จำนวนผู้เสียชีวิต จำนวนผู้บาดเจ็บ ประเภท และจำนวนของยานพาหนะที่เกิดเหตุด้วยการประเมินผลความถูกต้องของผลลัพธ์เปรียบเทียบกับต้นฉบับรายงานอุบัติเหตุ ดำเนินการโดยผู้วิจัยร่วมกับผู้เชี่ยวชาญด้านเทคโนโลยีสารสนเทศรวม 3 ท่าน เพื่อนำผลลัพธ์ความถูกต้องในแต่ละประเด็นมาเปรียบเทียบกัน และอภิปรายผลด้วยสถิติเชิงพรรณนา (Descriptive Statistic)

5.2 การประเมินความสามารถในการใช้งาน (Usability Testing) จากการทบทวนวรรณกรรมที่เกี่ยวข้อง ซึ่งรวบรวมจากหนังสือ ตำรา และเอกสารที่เกี่ยวข้องที่ได้รับการยอมรับในระดับนานาชาติ แบ่งเป็นเป้าหมายการใช้งานได้ที่กำหนดโดยไอเอสโอ [23] ได้กำหนดองค์ประกอบของการใช้งานได้ของระบบเทคโนโลยีสารสนเทศในมาตรฐานชุด ISO 9241-11 ซึ่งเป็นมาตรฐานทางการยศาสตร์ที่เกี่ยวกับการทำงานของมนุษย์ร่วมกับคอมพิวเตอร์ 3 ด้าน ดังนี้ คือ 1) ด้านประสิทธิผล (Effectiveness) คือความถูกต้องของเป้าหมายที่ผู้ใช้บรรลุในขอบเขตของระบบ 2) ด้านประสิทธิภาพ (Efficiency) คือ ปริมาณทรัพยากรที่ระบบใช้เพื่อให้บรรลุเป้าหมายอย่างสมบูรณ์ และ 3) ด้านความพึงพอใจ (Satisfaction) คือความสะดวกสบายและความยอมรับของผู้ใช้ในการใช้งานระบบ และเมื่อพิจารณาเป้าหมายการใช้งานได้ที่กำหนดโดยนักวิจัย ได้แก่ [24-29] จัดแบ่งประเภทเป้าหมายของการใช้งานได้เป็น 4 ด้าน คือ 1) ความมีประโยชน์ (Usefulness) ของระบบที่ช่วยผู้ใช้ให้บรรลุผลสำเร็จของงาน และทำงานได้ถูกต้อง 2) ประสิทธิภาพและประสิทธิภาพ (Effectiveness and Efficiency) คือความง่ายและความรวดเร็วในการทำงานให้บรรลุผลสำเร็จ ช่วยป้องกันข้อผิดพลาดในการทำงาน และการแก้ไขเมื่อเกิดข้อผิดพลาดขึ้น 3) การเรียนรู้ (Learnability) คือความง่ายในการเรียนรู้วิธีใช้งานระบบ หรือหากผู้ใช้ไม่ได้ใช้งานระบบระยะเวลาหนึ่ง เมื่อกลับมาใช้งานอีกครั้งยังคงจดจำวิธีใช้งานได้อย่างรวดเร็ว และ 4) ความพึงพอใจของผู้ใช้ (Likability) ที่ระบบสามารถทำให้ผู้ใช้เกิดความพึงพอใจในการใช้งานได้เพียงใด สอดคล้องกับ [30] ที่กำหนดเป้าหมายของการใช้งานในแนวทางเดียวกันแต่แยกการจดจำ (Memorability) ออกมาจากองค์ประกอบด้านการเรียนรู้ เพื่อประเมินว่าจะใช้เวลาเพียงใดเมื่อผู้ใช้กลับไปใช้งานอีกครั้ง หลังจากไม่ได้ใช้งานเป็นระยะเวลาหนึ่ง

เมื่อพิจารณาในรายละเอียดที่เกี่ยวข้องกับบริบทของผู้ใช้ โดย [31] แบ่งองค์ประกอบความสามารถในการใช้งานเป็น 1) ส่วนที่เป็นสมรรถนะของผู้ใช้ (Objective) เช่น ประสิทธิภาพของผู้ใช้งาน ประสิทธิภาพที่ได้จากการใช้งาน และความสามารถในการเรียนรู้ของผู้ใช้งาน และ 2) ส่วนที่เป็นความคิดเห็นส่วนตัวของผู้ใช้

(Subjective) เช่น ความยืดหยุ่นในการใช้งาน และความพึงพอใจของผู้ใช้ ผลการสังเคราะห์เอกสารที่เกี่ยวข้องนำมาเปรียบเทียบกับตารางที่ 1 จากนั้นกำหนดเป็นองค์ประกอบการประเมินความสามารถในการใช้งานในงานวิจัยชิ้นนี้

ตารางที่ 1 องค์ประกอบความสามารถในการใช้งาน

		[23]	[24]	[24]	[26]	[27]	[28]	[29]	[30]	In this Article
User	Usefulness							✓		
Performance	Efficiency	✓		✓		✓	✓	✓	✓	✓
(Objective)	Effectiveness	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Learnability		✓	✓	✓	✓	✓	✓	✓	✓
	Memorability			✓		✓	✓		✓	
User Opinion	Flexibility		✓		✓					✓
(Subjective)	Satisfaction	✓	✓			✓	✓		✓	✓

5.3 การกำหนดองค์ประกอบในการประเมินความสามารถในการนำเสนอสารสนเทศด้วยภาพข้อมูลผ่านเว็บไซต์ ที่เกี่ยวข้อง 5 องค์ประกอบ คือ 1) ด้านประสิทธิภาพ (Efficiency) เพื่อประเมินความเร็วในการปฏิสัมพันธ์ระหว่างผู้ใช้กับระบบค้นคืนสารสนเทศผ่านเว็บไซต์ในบริบทที่แตกต่างกัน 2) ด้านประสิทธิผล (Effectiveness) เพื่อประเมินความถูกต้องของผลลัพธ์ และกระบวนการได้มาซึ่งผลลัพธ์ 3) ด้านการเรียนรู้ของผู้ใช้ (Learnability) เพื่อประเมินความยากง่ายในการใช้งานที่ส่งผลต่อการเรียนรู้ของผู้ใช้ 4) ด้านความยืดหยุ่นในการใช้งาน (Flexibility) เพื่อประเมินความเหมาะสมในการแสดงผลบนฮาร์ดแวร์หรือซอฟต์แวร์ที่แตกต่างกัน และ 5) ด้านความพึงพอใจในการใช้งาน (Satisfaction) เพื่อประเมินความพึงพอใจในการใช้งาน และดึงดูดให้ผู้ใช้ใช้งานตั้งแต่ต้นจนจบกระบวนการ

5.4 การออกแบบเครื่องมือในการช่วยแบบสอบถามเพื่อประเมินความสามารถในการนำเสนอสารสนเทศด้วยภาพข้อมูลผ่านเว็บไซต์ จำนวน 20 ข้อ ผ่านการประเมินความเที่ยงตรงตามเนื้อหา (Content Validity) โดยพิจารณาจากค่าความสอดคล้องระหว่างข้อถามกับวัตถุประสงค์ (Index of Item Objective Congruence: IOC) จากผู้เชี่ยวชาญ 3 ท่าน ได้แก่ 1) ผู้เชี่ยวชาญด้านเทคโนโลยีสารสนเทศ 2) ผู้เชี่ยวชาญด้านข้อมูลจราจร และ 3) ผู้เชี่ยวชาญด้านสถิติและการวิจัย โดยแบ่งระดับการวัดคุณภาพตามมาตรวัดลิเคิร์ต [32] 5 ระดับ (5 -Point Likert Scale) กำหนดให้ (1) หมายถึง ระดับน้อยที่สุด (2) หมายถึง ระดับน้อย (3) หมายถึง ระดับปานกลาง (4) หมายถึง ระดับมาก และ (5) หมายถึง ระดับมากที่สุด

5.5 การเลือกประชากรและกลุ่มตัวอย่าง ในการประเมินความสามารถนำเสนอสารสนเทศด้วยภาพข้อมูลผ่านเว็บไซต์จากผู้ใช้งาน กำหนดประชากร คือ ผู้ที่สัญจรด้วยยานพาหนะทางถนนเป็นหลัก ซึ่งถือว่าทั้งหมดมีส่วนกับอุบัติเหตุทางถนนทั้งทางตรงและทางอ้อม ไม่ว่าจะเป็นผู้ขับขี่ก่อให้เกิดอุบัติเหตุ ผู้ที่มีส่วนได้ส่วนเสีย และผู้ได้รับผลกระทบ ตลอดจนมีส่วนร่วมในการป้องกันอุบัติเหตุ ด้วยการสุ่มตัวอย่างแบบไม่ใช้ความน่าจะเป็น (Non-probability Sampling) จากนั้นเลือกคัดกลุ่มตัวอย่างแบบเฉพาะเจาะจง (Purposive Sampling) [33] โดยพิจารณาจากการตัดสินใจของผู้วิจัยเอง เพื่อให้คุณลักษณะของกลุ่มที่เลือกเป็นไปตามวัตถุประสงค์ของการวิจัยที่ผู้ประเมินต้องอาศัยความรู้ ความชำนาญ และประสบการณ์ในเรื่องนั้น ๆ ตามแนวคิดการกำหนดจำนวนผู้ประเมินผลไว้ว่า 5 คน จะพบปัญหาในระบบได้ 85% และถ้าผู้ทดสอบ 15 คนหรือมากกว่าจะพบปัญหาทั้งหมดภายในระบบ [34]

ดังนั้นในงานวิจัยชิ้นนี้กำหนดผู้ประเมินทั้งสิ้น 30 คน มีระยะเวลาเก็บข้อมูลตั้งแต่เดือนตุลาคม ถึงเดือนพฤศจิกายน 2563 รวมทั้งสิ้น 2 เดือน

5.6 การวิเคราะห์ข้อมูลจากแบบสอบถามจำนวน 30 ชุด วิเคราะห์ข้อมูลด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูปทางสถิติในการวิเคราะห์สถิติเชิงพรรณนา เพื่อนำเสนอผลการประเมินความสามารถนำเสนอสารสนเทศด้วยภาพข้อมูลผ่านเว็บไซต์ ในรูปแบบตารางค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน โดยแปลผลค่าเฉลี่ยในแต่ละด้านว่าอยู่ในระดับใด ตามเกณฑ์ 5 ระดับ [35] กำหนดให้ 4.51 – 5.00 หมายถึง ระดับมากที่สุด, 3.51 – 4.50 หมายถึง ระดับมาก, 2.51 – 3.50 หมายถึง ระดับปานกลาง, 1.51 – 2.50 หมายถึง ระดับน้อย และ 1.00 – 1.50 หมายถึง ระดับน้อยที่สุด

5.7 การแยกผู้ใช้ออกเป็น 3 กลุ่ม กลุ่มละ 10 คน ตามบริบทการใช้งาน ได้แก่ 1) กลุ่มผู้เชี่ยวชาญด้านคอมพิวเตอร์ ที่ปฏิบัติงานหรือมีประสบการณ์ใช้งานคอมพิวเตอร์เฉพาะทางมากกว่า 10 ปีขึ้นไป 2) กลุ่มผู้ทำหน้าที่ข้อมูลจราจร ที่ปฏิบัติหน้าที่บังคับใช้กฎหมาย การวิจัย และเทคโนโลยีด้านจราจร หรือที่เกี่ยวข้อง และ 3) กลุ่มผู้ใช้ทั่วไป ที่เคยใช้งานระบบสารสนเทศด้านจราจรในลักษณะใดลักษณะหนึ่ง ด้วยสถิติวิเคราะห์เชิงอนุมาน (Inferential Statistic) โดยใช้สถิติทดสอบเอฟ (F-Test Statistic) สำหรับวิเคราะห์ความแปรปรวนทางเดียว ด้วยสถิติ One-way ANOVA แบบ LSD (Least Significant Difference Test) ที่มีลักษณะเชิงก้าวหน้ามากที่สุด (Most Liberal Test) [36] ของสถิติที่ใช้เพื่อการเปรียบเทียบค่าเฉลี่ยรายคู่ เนื่องจากมีความไวหรืออำนาจในการเปรียบเทียบสูง เพื่อหาโอกาสที่จะพบความแตกต่างระหว่างค่าเฉลี่ยผลการทดสอบความสามารถในการใช้งาน ทั้ง 5 องค์ประกอบของผู้ใช้ทั้ง 3 กลุ่มที่แตกต่างกันนั้น มีปฏิสัมพันธ์กันหรือไม่ และเป็นไปในทิศทางเดียวกันหรือไม่

ผลการวิจัย

งานวิจัยชิ้นนี้เป็นงานวิจัยเชิงประยุกต์ มีวัตถุประสงค์เพื่อออกแบบและพัฒนากระบวนการสกัดข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่ที่รวบรวมมาจากเว็บไซต์ www.thairsc.com ด้วยเทคนิคการขูดเว็บ และประเมินความสามารถนำเสนอสารสนเทศด้วยภาพข้อมูลผ่านเว็บไซต์ โดยมีผลการวิจัยดังนี้

1) ผลการออกแบบและพัฒนากระบวนการสกัดข้อมูล เริ่มจากผู้วิจัยนำเข้าข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่ที่เข้าสู่ฐานข้อมูล จากนั้นจะคิวรี่ส่วนหัวของรายงาน และส่วนรายละเอียดของรายงาน เพื่อสกัดข้อมูลตามกระบวนการที่ได้ออกแบบและพัฒนาไว้ในลักษณะวนลูป ใน 6 ประเด็น ได้แก่ จังหวัด วันที่ จำนวนผู้เสียชีวิต จำนวนผู้บาดเจ็บ ประเภทและจำนวนของยานพาหนะ ดังภาพที่ 5

Code	#ContentPlace Holder1_lb_title	#section .clearfix	Extract Data
003013	Accident 2 Cars Collided, 3 Deaths, 5 Injuries, Si Kiew District, Nakhon Ratchasima (15-03-2020, 04.50 a.m.)	Accident: 2 Cars Collided 3 Deaths 5 Injuries Si Kiew District Nakhon Ratchasima (15032020 04.50 a.m.)Accident Notification CentreRoad Accident Victims Protection Co. Ltd. Department of Disaster Prevention and MitigationAccident ReportNotification No: EAccident : 63/001/0004402 Source: JS100. Date of Claim: Sunday March 15th 2020 at about 08.00 a.m. Date of Accident: Sunday March 15th 2020 at about 04.50 a.m.Accident Point: Highway 2 Mitrphap Road KM.104105 Utum bridge are Mitrphap Si Kiew District Nakhon Ratchasima. Accident Condition: straightway Weather Condition: rain. Responsibility area: Si Kiew Police Station Tel. 044411010Enquiry officer/Case owners: Tel. Accident scene: Accident: 2 cars collided.Photo by: Motor Vehicle(s) involved in the accident1. Sedan car Honda Civic grey color license plate registration no. xxxxx Maha Sarakham.2. Pickup car Mitsubishi white color license plate registration no. xxxxxx Udon Thani.List of deaths: 3 Deaths (Male:2 Female: 1) List of injured: 5 Injuries (Male: 3 Female: 2) taken to Si Kiew Hospital. Injuries Identification: Cause of the accident: (The real causes of the accident is being investigated.)Insurance Assistance and coordination1. License plate registration no. xxxxx Maha Sarakham compulsory insurance is NOT FOUND 2. License plate registration no. xxxxxx Udon Thani compulsory insurance is NOT FOUND Hotline(s) Si Kiew Police Station (Tel. 044411010) Promtham Rescue Si Kiew District (Tel. 044984115) Si Kiew Hospital (Tel. 044986240) Department of Disaster Prevention and Mitigation (Call Center 1784) Road Accident Victims Protection Company Limited IOC and the Office of Insurance Commission (OIC)Road Accident Victims Protection Co. Ltd.44/1 Rungrojthanakul Building 11th Floor Ratchadapisek Rd. Huaykwang Bangkok 10310Call Center 1791 Tel: 021009191 Fax: 0264302934www.rvp.co.th www.thairsc.com	Province: Nakhon Ratchasima Date: 2020-03-15 Motor Vehicle(s) : [1] CAR [1] PICK UP Death(s) 3 Injured(s) 5

ภาพที่ 5 ผลลัพธ์จากการออกแบบและพัฒนากระบวนการสกัดข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่

การนำเสนอสารสนเทศด้วยภาพข้อมูลผ่านเว็บไซต์จะใช้งานผ่านเว็บเบราว์เซอร์ (Web Browser) เบื้องต้นผู้ใช้จะต้องระบุเงื่อนไขวันเริ่มต้น วันสิ้นสุด และเลือกจังหวัดที่ต้องการแสดงผลในแต่ละครั้ง หรือหากผู้ใช้ต้องการค้นคืนสารสนเทศที่มีความละเอียดมากขึ้น ก็สามารถระบุเงื่อนไขชนิดและจำนวนยานพาหนะในที่เกิดเหตุ ตลอดจนจำนวนผู้บาดเจ็บและจำนวนผู้เสียชีวิตร่วมด้วยได้ ดังภาพที่ 6

str_date : 2019-01-01 end_date : 2020-12-31

ผู้ขับขี่ > 2 ผู้เสียชีวิต > 2

Advanced Search

รถยนต์ >= 1 จักรยาน/รถจักรยานยนต์ >= 1

Northern

- Chiang Mai
- Chiang Rai
- Lamphang
- Lamphun
- Mae Hong Son
- Nan
- Phayao
- Phrae
- Uttaradit

Northeastern

- Amnat Charoen
- Bueng Kan
- Buriram
- Chaiyaphum
- Kalasin
- Khon Kaen
- Loei
- Maha Sarakham
- Mukdahan
- Nakhon Phanom
- Nakhon Ratchasima
- Nong Bus Lamphu
- Nong Khai
- Roi Et
- Sakon Nakhon
- Sisaket
- Surin
- Ubon Ratchathani
- Udorn Thani
- Yasothon

Central

- Ang Thong
- Bangkok
- Chai Nat
- Kamphaeng Phet
- Kalasin
- Lopburi
- Nakhon Nayok
- Nakhon Pathom
- Nakhon Sawan
- Nonthaburi
- Pathum Thani
- Phetchabun
- Phichit
- Phitsanulok
- Phra Nakhon Si Ayutthaya
- Samut Prakan
- Samut Sakhon
- Samut Songkhram
- Saraburi
- Sing Buri
- Sukhothai
- Suphan Buri
- Uthai Thani

Eastern

- Chachoengsao
- Chanthaburi
- Chonburi
- Prachinburi
- Rayong
- Sa Kaeo
- Trat

Western

- Kanchanaburi
- Phetchaburi
- Prachuap Khiri Khan
- Ratchaburi
- Tak

Southern

- Chumphon
- Krabi
- Nakhon Si Thammarat
- Narathiwat
- Pattani
- Phang Nga
- Phatthalung
- Phuket
- Ranong
- Satun
- Songkhla
- Surat Thani
- Trang
- Yala

Show By: จำนวนการเกิดอุบัติเหตุ

จำนวน (ครั้ง)	เสียชีวิต (คน)	บาดเจ็บ (คน)
57	114	143

Motor Vehicle(s) involved in the accident

proceed time : 0.5523 วินาที

memory usage : 10.96 mb

ลำดับ	ส่วนหัวของรายงาน	ส่วนรายละเอียดของรายงาน	ผลการวิเคราะห์ข้อมูล
002013	Accident 2 Cars Collided, 3 Deaths, 5 Injuries, Si Kiew District, Nakhon Ratchasima (15-03-2020, 04:50 a.m.)	Accident: 2 Cars Collided 3 Deaths 5 Injuries Si Kiew District Nakhon Ratchasima (15032020 04:50 a.m.) Accident Notification Centre/Road Accident Victims Protection Co. Ltd. Department of Disaster Prevention and Mitigation/Accident Report/Notification No: E/Accident : 63/001/0004402 Source: JS100. Date of Claim: Sunday March 15th 2020 at about 06:00 a.m. Date of Accident: Sunday March 15th 2020 at about 04:50 a.m. Accident Point: Highway 2 Mitphap Road Kh.104105 Uturn bridge area Mitphap Si Kiew District Nakhon Ratchasima. Accident Condition: straightway Weather Condition: rain. Responsibility area: Si Kiew Police Station Tel. 044411010/Enquiry officer/Case owners: Tel. Accident scene: Accident: 2 cars collided. Photo by: Motor Vehicle(s) involved in the accident: 1. Sedan car Honda Civic grey color license plate registration no. xxxxxx Maha Sarakham. 2. Pickup car Mitsubishi white color license plate registration no. xxxxxx Udon Thani. List of deaths: 2 Deaths (Male/2 Female: 1) List of injured: 5 Injuries (Male: 3 Female: 2) taken to Si Kiew Hospital. Injuries Identification: Cause of the accident: (The real causes of the accident is being investigated). Insurance Assistance and coordination: License plate registration no. xxxxxx Maha Sarakham compulsory insurance is NOT FOUND 2. License plate registration no. xxxxxx Udon Thani compulsory insurance is NOT FOUND Hotline(s) Si Kiew Police Station (Tel. 044411010) Promtham Rescue Si Kiew District (Tel. 044964115) Si Kiew Hospital (Tel. 044962640) Department of Disaster Prevention and Mitigation (Call Center 1794) Road Accident Victims Protection Company Limited IOC and the Office of Insurance Commission (OIC) Road Accident Victims Protection Co. Ltd.44/1 Rungrojthanakul Building 11th Floor Ratchadasek Rd. Huaykwang Bangkok 10310/Call Center 1791 Tel: 021009191 Fax: 026430294/www.rvp.co.th www.thaifsc.com	Province: Nakhon Ratchasima Date: 2020-03-15 Motor Vehicle(s) : [1] CAR [1] PICK UP Death(s) 3 Injured(s) 5

Province	Accident(s)	Death(s)	Injury(s)	CAR	PICK UP	MOTORCYCLE	BICYCLE	TRUCK	VAN	BUS	Pedestrian
Nakhon Phanom	4	9	5	5	2	0	0	0	0	0	0
Nakhon Ratchasima	23	51	71	21	14	8	0	7	1	7	0
Nakhon Nayok	3	6	1	3	1	1	0	1	0	1	0
Nakhon Pathom	11	19	23	9	5	7	0	5	1	5	0
Nakhon Sawan	7	12	24	9	4	2	0	1	0	1	0
Nakhon Si Thammarat	9	17	19	8	8	7	0	0	0	0	0

ภาพที่ 6 การนำเสนอสารสนเทศรายงานอุบัติเหตุทางถนนรายใหญ่ด้วยภาพข้อมูลผ่านเว็บไซต์

จากภาพที่ 6 การนำเสนอสารสนเทศด้วยภาพข้อมูลผ่านเว็บไซต์ที่มีปฏิสัมพันธ์กับผู้ใช้แบบแอกทีฟด้วยการนำเมาส์ไปวางบนตามตำแหน่งที่ต้องการ แบ่งเป็นสถิติการเกิดอุบัติเหตุเชิงพื้นที่ด้วยกราฟแผนที่ประเทศไทย เพื่อเปรียบเทียบสถิติภาพรวมรายจังหวัด โดยไล่ความเข้มของสีจากสีขาว สีเหลือง สีส้ม และสีแดงในการแสดงผลตามปริมาณความถี่การเกิดอุบัติเหตุ ซึ่งสามารถเลือกแสดงผลได้ว่าต้องการแสดงผลด้วยจำนวนการเกิด

อุบัติเหตุ จำนวนผู้บาดเจ็บ หรือจำนวนผู้เสียชีวิต ร่วมกับเปรียบเทียบจำนวนการเกิดอุบัติเหตุตามประเภทของยานพาหนะด้วยแผนภูมิวงกลม ตลอดจนการนำเสนอสารสนเทศด้วยตารางที่สามารถเรียงลำดับข้อมูล (Sort) ในแต่ละคอลัมน์ที่สนใจได้ด้วยตนเอง จากมากที่สุดไปยังน้อยที่สุด หรือน้อยที่สุดไปยังมากที่สุด

2) การประเมินความถูกต้องของผลลัพธ์จากกระบวนการสกัดข้อมูลใน 6 ประเด็น ได้แก่ จังหวัด วันที่ จำนวนผู้เสียชีวิต จำนวนผู้บาดเจ็บ ประเภทและจำนวนของยานพาหนะ ประเมินผลโดยผู้วิจัยร่วมกับผู้เชี่ยวชาญด้านเทคโนโลยีสารสนเทศรวม 3 ท่าน ด้วยการเลือกเงื่อนไขหลักเป็นจังหวัดที่เกิดอุบัติเหตุที่ละจังหวัด จำนวน 77 ครั้ง ตามจำนวนจังหวัดของประเทศไทย ร่วมกับเงื่อนไขการแสดงผลอื่นๆ (ถ้ามี) เป็นเงื่อนไขรองแบบสุ่ม แล้วประเมินความถูกต้องของผลลัพธ์แต่ละรายการว่าถูกต้องหรือไม่ หากไม่ปรากฏผลลัพธ์การนำเสนอสารสนเทศด้วยภาพข้อมูล เนื่องจากผลลัพธ์ไม่ตรงกับเงื่อนไขรองใดเลย จะต้องทำการเลือกเงื่อนไขรองใหม่ กำหนดผลการประเมินเป็นถูกต้องได้ 1 คะแนน และไม่ถูกต้องได้ 0 คะแนน เช่นเดียวกับผลลัพธ์ในข้อใดไม่ชัดเจน ให้ถือว่าไม่ถูกต้อง โดยแต่ละรายการหากมีคะแนนการประเมินมากกว่า 2 ใน 3 จะถือว่าผลลัพธ์นั้นถูกต้อง เมื่อครบรอบจะทำการประเมินอีกครั้งจนครบ 3 รอบ แบบเงื่อนไขหลักคงที่ และเงื่อนไขรองเป็นอิสระต่อกัน แล้วนำผลการประเมินมาหาค่าเฉลี่ยเพื่อให้ค่าความคลาดเคลื่อนอยู่ในระดับที่ยอมรับได้ ผลการประเมินดังตาราง 2

กำหนดการแปลผลเป็นร้อยละ 5 ระดับ เพื่อให้ง่ายต่อการทำความเข้าใจ โดยตัดแปลงจากวิธีการใช้ร้อยละของคะแนนในการกำหนดช่วง [37] ดังนี้ ต่ำกว่าร้อยละ 50 หมายถึง ต้องปรับปรุง, ร้อยละ 50-59 หมายถึง พอใช้, ร้อยละ 60-69 หมายถึง ปานกลาง, ร้อยละ 70-79 หมายถึง ดี และ ร้อยละ 80-100 หมายถึง ดีมาก

ตารางที่ 2 ผลการประเมินความถูกต้องของผลลัพธ์จากกระบวนการสกัดข้อมูล

ผลลัพธ์จากกระบวนการสกัดข้อมูล	ผลการประเมินความถูกต้องของผลลัพธ์จากกระบวนการสกัดข้อมูล							
	ครั้งที่ 1 (N=411)		ครั้งที่ 2 (N=442)		ครั้งที่ 3 (N=474)		ค่าเฉลี่ย	
	ผลลัพธ์ที่ถูกต้อง	ร้อยละ	ผลลัพธ์ที่ถูกต้อง	ร้อยละ	ผลลัพธ์ที่ถูกต้อง	ร้อยละ	ร้อยละ	แปลผล
1) จังหวัด	403	98.05	434	98.19	466	98.31	98.19	ดีมาก
2) วันที่	400	97.32	438	99.10	452	95.36	97.26	ดีมาก
3) จำนวนผู้เสียชีวิต	322	78.35	342	77.38	324	68.35	74.69	ดี
4) จำนวนผู้บาดเจ็บ	296	72.02	285	64.48	309	65.19	67.23	ปานกลาง
5) ประเภทของยานพาหนะ	398	96.84	431	97.51	464	97.89	97.41	ดีมาก
6) จำนวนของยานพาหนะ	322	78.35	331	74.89	388	81.86	78.36	ดี

จากตารางที่ 2 พบว่า ผลการประเมินความถูกต้องของผลลัพธ์จากกระบวนการสกัดข้อมูล ในประเด็นจังหวัด วันที่ และประเภทของยานพาหนะ อยู่ในระดับดีมาก (ค่าเฉลี่ย ร้อยละ 98.19, 97.26 และ 97.41 ตามลำดับ) แสดงถึง กระบวนการสกัดข้อมูลด้วยการสร้างคลังคำศัพท์ร่วมกับการรู้จำเอนทิตีให้ผลลัพธ์ความถูกต้องในระดับสูง แต่การสร้างคลังคำศัพท์เหมาะสมกับข้อมูลที่กำหนดขอบเขตได้แน่นอน ดังนั้นการรวบรวมข้อมูลคำศัพท์ต้องมีกระบวนการที่น่าเชื่อถือ เช่น ประเภทยานพาหนะ หรือรายชื่อจังหวัดที่มักเขียนผิด เป็นต้น นอกจากนี้ในประเด็นของจำนวนผู้เสียชีวิต และจำนวนของยานพาหนะอยู่ในระดับดี (ค่าเฉลี่ย ร้อยละ 74.69 และ

78.36 ตามลำดับ) แสดงถึงการกำหนดการกำหนดรูปแบบอักษรร่วมกับการรู้จำเอนทิตี ยังให้ผลลัพธ์การสกัดข้อมูลที่น่าพอใจ แต่อย่างไรก็ดี ผลการประเมินความถูกต้องของจำนวนผู้บาดเจ็บ ซึ่งใช้กระบวนการสกัดข้อมูลเกี่ยวกับการสกัดข้อมูลจำนวนผู้เสียชีวิต แต่มีค่าเฉลี่ยความถูกต้องต่ำที่สุด จากการตรวจสอบพบว่าเงื่อนไขการค้นหาข้อมูลตามลำดับก่อนหลังมีผลต่อความถูกต้องของการสกัดข้อมูล เนื่องจากรูปแบบการบันทึกข้อมูลที่ไม่เป็นมาตรฐาน เช่น การใส่เครื่องหมาย – เมื่อไม่มีจำนวนผู้เสียชีวิตหรือจำนวนผู้บาดเจ็บ แทนการใส่ตัวเลข 0 เป็นต้น ตลอดจนการใส่รายละเอียดอื่นร่วมกับตัวเลข เช่น 5 (Male:3 Female:2) เป็นต้น

3) ผลประเมินความสามารถในการนำเสนอสารสนเทศรายงานอุบัติเหตุทางถนนรายใหญ่ด้วยภาพข้อมูลผ่านเว็บไซต์ จากกลุ่มตัวอย่าง 30 คน โดยให้ผู้ใช้แต่ละคนกำหนดเงื่อนไขในการนำเสนออย่างอิสระ แล้วประเมินผลผ่านแบบสอบถามออนไลน์คนละ 3 ครั้ง ที่แสดงเฉพาะประเด็นข้อคำถามโดยไม่แสดงว่าแต่ละข้อคำถามอยู่ในองค์ประกอบการใช้งานด้านใด และในแต่ละครั้งจะทำการสลับข้อคำถาม จากนั้นจะนำผลการประเมินมาหาค่าเฉลี่ย เพื่อลดค่าความคลาดเคลื่อนเชิงสถิติ (Statistical Error) [38] ให้อยู่ในระดับที่ยอมรับได้ ในงานวิจัยชิ้นนี้แบ่งเป็น 5 องค์ประกอบ ได้แก่ ด้านประสิทธิภาพ ด้านประสิทธิผล ด้านความสามารถในการเรียนรู้ ด้านความยืดหยุ่น และด้านความพึงพอใจโดยข้อมูลผู้ใช้งานมีรายละเอียดดังตารางที่ 3

ตารางที่ 3 ผลการประเมินความสามารถในการนำเสนอสารสนเทศรายงานอุบัติเหตุทางถนนรายใหญ่ด้วยภาพข้อมูลผ่านเว็บไซต์

ข้อ	ประเด็นคำถามการใช้งาน	\bar{X}	S.D.	แปลผล
ด้านประสิทธิภาพ (Efficiency)				
1.	ความรวดเร็วในการนำเสนอผลการสกัดข้อมูล	3.23	0.74	ปานกลาง
2.	ความรวดเร็วในการนำเสนอภาพข้อมูล	3.57	0.67	ดี
3.	ระยะเวลาการตอบสนองของภาพข้อมูลเมื่อถูกกระทำ	3.73	0.60	ดี
	ค่าเฉลี่ย	3.51	0.64	ดี
ด้านประสิทธิผล (Effectiveness)				
4.	ความถูกต้องของการแสดงผลข้อมูลตามเงื่อนไขการค้นหา	3.37	0.78	ปานกลาง
5.	ความถูกต้องของผลการสกัดข้อมูลรายงานอุบัติเหตุ	3.30	0.79	ปานกลาง
6.	ความถูกต้องของข้อมูลเมื่อนำเสนอในรูปแบบภาพข้อมูล	3.37	0.70	ปานกลาง
7.	ความถูกต้องของการจัดวางรายชื่อจังหวัดเพื่อการค้นหา	3.37	0.70	ปานกลาง
8.	ความถูกต้องของการใช้สีในการแสดงความแตกต่างของข้อมูลแต่ละประเภท	3.77	0.70	ดี
	ค่าเฉลี่ย	3.43	0.73	ปานกลาง
ด้านความสามารถในการเรียนรู้ (Learnability)				
9.	สามารถเรียนรู้การใช้ระบบได้เอง โดยไม่ต้องใช้ผู้เชี่ยวชาญช่วยเหลือ	3.93	0.53	ดี
10.	สามารถใช้งานระบบได้อย่างรวดเร็ว	4.00	0.49	ดี
11.	สามารถเข้าใจกับข้อมูลที่นำเสนอในลักษณะภาพข้อมูลได้ง่าย	4.33	0.41	ดี
12.	สามารถทำความเข้าใจกับกระบวนการได้มาซึ่งผลลัพธ์ได้ด้วยตนเอง	3.87	0.60	ดี
	ค่าเฉลี่ย	4.03	0.52	ดี

ตารางที่ 3 (ต่อ)

ข้อ	ประเด็นคำถามการใช้งาน	\bar{X}	S.D.	แปลผล
ด้านความยืดหยุ่น (Flexibility)				
13.	การนำเสนอภาพข้อมูลมีการตอบสนองและปรับเปลี่ยนได้ตามที่ผู้ใช้งานต้องการ	3.77	0.64	ดี
14.	ความเหมาะสมของหน้าเว็บไซต์เมื่อแสดงผลบนอุปกรณ์ที่มีขนาดต่างกัน	3.80	0.73	ดี
15.	ความเหมาะสมของหน้าเว็บไซต์เมื่อแสดงผลบนเว็บเบราว์เซอร์ที่ต่างกัน	3.23	0.67	ปานกลาง
	ค่าเฉลี่ย	3.60	0.70	ดี
ด้านความพึงพอใจ (Satisfaction)				
16.	ระบบมีรูปแบบการจัดวางองค์ประกอบที่ชัดเจน	4.13	0.48	ดี
17.	สีสันท่านำมาใช้แสดงผลมีความโดดเด่น สวยงาม	3.90	0.46	ดี
18.	ขนาดของรูปภาพ และข้อความ มีความเหมาะสม	3.93	0.53	ดี
19.	ความเหมาะสมของการเลือกใช้ภาพข้อมูลแต่ละประเภทในการนำเสนอ	4.13	0.56	ดี
20.	ความเหมาะสมของการเลือกใช้สีในการแสดงระดับความรุนแรงของอุบัติเหตุ	3.90	0.58	ดี
	ค่าเฉลี่ย	4.00	0.52	ดี
	ค่าเฉลี่ยรวมทุกองค์ประกอบ	3.73	0.63	ดี

จากตารางที่ 3 ผลการประเมินสรุปได้ว่าการนำเสนอสารสนเทศรายงานอุบัติเหตุทางถนนรายใหญ่ด้วยภาพข้อมูลผ่านเว็บไซต์มีความสามารถในการใช้งานอยู่ในระดับดี ($\bar{X} = 3.73$) เมื่อพิจารณาแต่ละด้าน พบว่า ผลการประเมินทุกด้านมีค่าเฉลี่ยอยู่ในระดับมากขึ้นไป โดยองค์ประกอบที่เป็นจุดเด่น คือ ด้านความสามารถในการเรียนรู้ และด้านความพึงพอใจ ซึ่งมีค่าเฉลี่ยที่ใกล้เคียงกัน ($\bar{X} = 4.03$ และ 4.00 ตามลำดับ) อย่างไรก็ตาม ระบบยังมีปัญหาในด้านความถูกต้องของข้อมูล อันจะเห็นผลการประเมินด้านประสิทธิผลที่มีค่าเฉลี่ยในระดับปานกลาง ($\bar{X} = 3.43$) ซึ่งเป็นค่าเฉลี่ยที่ต่ำที่สุดจากทุกองค์ประกอบ และเป็นที่น่าสังเกตว่าผลการประเมินความถูกต้องของการแสดงผลข้อมูลตามเงื่อนไขการค้นหา กับความถูกต้องของผลการสกัดข้อมูลรายงานอุบัติเหตุ อันมีที่มาจากระบบการสกัดข้อมูลในขั้นตอนเดียวกัน ผลการประเมินแตกต่างกันเล็กน้อย ($\bar{X} = 3.37$ และ 3.30 ตามลำดับ) แต่การกระจายตัวของข้อมูลกลับสูงกว่าประเด็นคำถามอื่น ๆ (S.D. = 0.78 และ 0.79 ตามลำดับ)

จากประเด็นดังกล่าว ผู้วิจัยจึงได้แบ่งผู้ใช้เป็น 3 กลุ่มตามบริบทในการใช้งาน กลุ่มละ 10 คน ได้แก่ กลุ่มผู้เชี่ยวชาญด้านคอมพิวเตอร์ กลุ่มผู้ทำหน้าที่ข้อมูลจราจร และกลุ่มผู้ใช้ทั่วไป เพื่อวิเคราะห์ความสัมพันธ์ของผลการประเมินจากผู้ในแต่ละกลุ่มเปรียบเทียบหลังการประเมินรวมด้วยการวิเคราะห์ความแปรปรวนทางเดียว สำหรับการเปรียบเทียบไม่เกิน 3 กลุ่มในครั้งเดียวด้วยสถิติทดสอบเอฟ แบบ LSD รายละเอียดดังตารางที่ 4

ตารางที่ 4 ผลการประเมินความสามารถในการนำเสนอสารสนเทศรายงานอุบัติเหตุทางถนนรายใหญ่ด้วยภาพข้อมูลผ่านเว็บไซต์ แยกตามกลุ่มผู้ใช้

องค์ประกอบ	1	2	.sig	1	3	.sig	2	3	.sig
	\bar{X}	\bar{X}		\bar{X}	\bar{X}		\bar{X}	\bar{X}	
1. ด้านประสิทธิภาพ	3.37	3.57	.430	3.37	3.60	.442	3.57	3.60	.952
2. ด้านประสิทธิผล	3.48	3.34	.592	3.48	3.48	.908	3.34	3.48	.663
3. ความสามารถในการเรียนรู้	3.80	4.30	.020*	3.80	4.00	.133	4.30	4.00	.316
4. ด้านความยืดหยุ่น	3.63	3.40	.562	3.63	3.77	.893	3.40	3.77	.471
5. ด้านความพึงพอใจ	3.92	3.94	.929	3.92	4.14	.155	3.94	4.14	.194
ค่าเฉลี่ยทุกด้าน	3.66	3.73	.396	3.66	3.81	.106	3.73	3.81	.462

* นัยสำคัญทางสถิติที่ 0.05

** กลุ่ม 1 กลุ่มผู้เชี่ยวชาญด้านคอมพิวเตอร์; กลุ่ม 2 กลุ่มผู้ทำหน้าที่ข้อมูลจราจร; กลุ่ม 3 กลุ่มผู้ใช้ทั่วไป

จากตารางที่ 4 ผลการประเมินจากผู้ใช้งาน 3 กลุ่ม ได้แก่ กลุ่มที่ 1 กลุ่มผู้เชี่ยวชาญด้านคอมพิวเตอร์ กลุ่มที่ 2 กลุ่มผู้ทำหน้าที่ข้อมูลจราจร และกลุ่มที่ 3 กลุ่มผู้ใช้ทั่วไปมีค่าเฉลี่ยโดยรวมอยู่ในระดับดี ($\bar{X} = 3.66, 3.73$ และ 3.81 ตามลำดับ) เมื่อพิจารณาผลการประเมินในแต่ละด้าน พบว่า กลุ่มผู้เชี่ยวชาญด้านคอมพิวเตอร์ และกลุ่มผู้ใช้ทั่วไป มีผลการประเมินด้านความพึงพอใจสูงที่สุด ($\bar{X} = 3.92$ และ 4.14 ตามลำดับ) ในขณะที่กลุ่มผู้ทำหน้าที่ข้อมูลจราจร มีผลการประเมินด้านความสามารถในการเรียนรู้สูงที่สุด ($\bar{X} = 4.30$) โดยผลการประเมินดังกล่าวมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ (.sig) กับกลุ่มผู้เชี่ยวชาญด้านคอมพิวเตอร์ที่ระดับ 0.05 แสดงให้เห็นถึงประสบการณ์ในการใช้งานระบบด้านการจราจรที่เกี่ยวข้องส่งผลต่อความสามารถในการใช้งานระบบที่พัฒนาขึ้นมากกว่าประสบการณ์ใช้งานคอมพิวเตอร์และสมาร์ทโฟน

สรุปและอภิปรายผล

งานวิจัยชิ้นนี้ออกแบบและพัฒนากระบวนการสกัดข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่ โดยใช้วงจรการพัฒนาแบบรูปแบบนำตกแบบย้อนกลับได้เป็นกรอบในการดำเนินงาน รวบรวมข้อมูลเชิงคุณภาพที่อยู่ในลักษณะกึ่งมีโครงสร้างจากเว็บไซต์ www.thairsc.com ด้วยเทคนิคการขุดเว็บ เพื่อนำมาสกัดข้อมูลด้วยการสร้างคลังคำศัพท์ร่วมกับการรู้จำเอนทิตีใน 6 ประเด็น ได้แก่ จังหวัด วันที่ จำนวนผู้เสียชีวิต จำนวนผู้บาดเจ็บ ประเภท และจำนวนของยานพาหนะที่เกิดเหตุนำเสนอแก่ผู้ใช้ในรูปแบบภาพข้อมูลด้วย Google Charts ที่ทำงานส่งข้อมูลจากเครื่องลูกข่ายไปประมวลผลยังเครื่องแม่ข่ายก่อนจะนำเสนอสารสนเทศแก่ผู้ใช้ เพื่อเปรียบเทียบสถิติการเกิดอุบัติเหตุในแต่ละมิติ การประเมินความถูกต้องของผลลัพธ์จากกระบวนการสกัดข้อมูลจากผู้เชี่ยวชาญ 3 ท่าน กำหนดเงื่อนไขหลักคือรายชื่อจังหวัดร่วมกับเงื่อนไขรองอื่น ๆ พบว่า การสกัดข้อมูลชื่อจังหวัด และประเภทของยานพาหนะด้วยการสร้างคลังคำศัพท์ร่วมกับการรู้จำเอนทิตีมีความถูกต้องสูง (ค่าเฉลี่ย ร้อยละ 98.19 และ 97.41 ตามลำดับ) ระดับความถูกต้องดังกล่าวเป็นไปได้ทางเดียวกับกับงานวิจัยที่ใช้การรู้จำเอนทิตีจากข้อมูลที่ระบุไว้ในพจนานุกรม เพื่อสกัดชื่อภัตตาคารผู้สร้างวัดในสมัยล้านนา [11] มีความถูกต้องถึงร้อยละ 97 เช่นเดียวกับประยุกต์การรู้จำเอนทิตี เพื่อแยกข้อมูลภาษาไทย เช่น ชื่อบุคคล ชื่อองค์กร สถานที่ วันที่และเวลา เป็นต้น ที่กำหนดไว้ในพจนานุกรม ร่วมกับการจับคู่รูปแบบตามที่กำหนด [7] พบว่า มีความถูกต้องประมาณ 68-100% ตลอดจนการรวบรวมข้อมูลการท่องเที่ยวไทย [12] มาสร้างคลังคำศัพท์ เพื่อสกัดข้อมูล เช่น ร้านอาหาร โรงแรม การเลือกซื้อสินค้า และการท่องเที่ยว เป็นต้น ด้วยการใช้การรู้จำเอนทิตี ที่มีประสิทธิภาพมากกว่าร้อยละ 90 ดังนั้น

กระบวนการสกัดข้อมูลด้วยการสร้างคลังคำศัพท์ร่วมกับการรู้จำเอนทิตีที่มีประสิทธิภาพสูง แต่เหมาะสมกับข้อมูลที่มีความเฉพาะเจาะจงสูง และขึ้นอยู่กับความครบถ้วนและถูกต้องของคลังคำศัพท์

ผลการประเมินความสามารถในการนำเสนอสารสนเทศด้วยภาพข้อมูลผ่านเว็บไซต์ จากผู้ใช้ 30 คน ด้วยการสุ่มตัวอย่างโดยไม่ใช้ความน่าจะเป็นจากการเลือกกลุ่มตัวอย่างแบบเฉพาะเจาะจง โดยให้ผู้ใช้แต่ละคนทดลองใช้งานแล้วประเมินผลผ่านแบบสอบถามคนละ 3 ครั้งแล้วนำมาหาค่าเฉลี่ย เพื่อลดความคลาดเคลื่อนให้อยู่ในระดับที่ยอมรับได้ ผลการประเมินภาพรวมอยู่ในระดับดี แต่ระบบยังมีปัญหาด้านความถูกต้องของข้อมูล เนื่องจากความครอบคลุมของคลังคำศัพท์ และรูปแบบการบันทึกข้อมูลที่ไม่เป็นมาตรฐาน เมื่อแบ่งผู้ใช้เป็น 3 กลุ่มตามบริบทการใช้งานได้แก่ กลุ่มผู้เชี่ยวชาญด้านคอมพิวเตอร์ กลุ่มผู้ทำหน้าที่ข้อมูลจราจร และกลุ่มผู้ใช้ทั่วไป แล้ววิเคราะห์ด้วยสถิติทดสอบเอฟ แบบ LSD พบว่า กลุ่มผู้เชี่ยวชาญด้านคอมพิวเตอร์ และกลุ่มผู้ทำหน้าที่ข้อมูลจราจร มีความแตกต่างด้านการเรียนรู้อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 แสดงถึงประสิทธิภาพในการใช้งานระบบด้านการจราจรที่เกี่ยวข้องส่งผลต่อการเรียนรู้มากกว่าประสิทธิภาพใช้งานคอมพิวเตอร์และสมาร์ทโฟน ที่แม้จะมีประสิทธิภาพในการใช้งานคอมพิวเตอร์และสมาร์ทโฟนมากกว่า แต่ผลการประเมินด้านความสามารถในการเรียนรู้ต่ำกว่า เช่นเดียวกับ [39] ผลการวิจัยที่พบว่าประสิทธิภาพในการใช้งานคอมพิวเตอร์และสมาร์ทโฟนเป็นหนึ่งในตัวแปรสำคัญที่มีผลต่อความสามารถในการใช้งานเซทของที่นำเสนอสารสนเทศในรูปแบบต่าง ๆ ตามที่กำหนดให้กับผู้ใช้ผ่านสมาร์ทโฟน สอดคล้องกับผลการวิจัยในกลุ่มผู้สูงอายุ [40] พบว่า ประสิทธิภาพของผู้ใช้สมาร์ทโฟนส่งผลต่อความสามารถในการใช้งานของขนาดและรูปร่างของปุ่มกด

เห็นได้ว่าประสิทธิภาพในการใช้งานเป็นปัจจัยสำคัญที่ส่งผลต่อความสามารถในการใช้งาน ดังนั้นทุกระบบที่พัฒนาขึ้นจึงควรคำนึงถึงประสิทธิภาพของผู้ใช้งาน (User eXperience: UX) [41] ที่มีความสัมพันธ์ต่อการใช้งานของผู้ใช้งาน (Usability) และการเข้าถึง (Accessibility) ที่สร้างปฏิสัมพันธ์ของผู้ใช้งานต่อการใช้งานระบบงาน ผ่านส่วนต่อปฏิสัมพันธ์กับผู้ใช้ (User Interface: UI) ประเด็นดังกล่าวจึงสอดคล้องกับข้อค้นพบในงานวิจัยชิ้นนี้ที่ประสิทธิภาพในการใช้งาน ส่งผลโดยตรงต่อความสามารถในการใช้งาน โดยองค์ประกอบด้านความสามารถในการเรียนรู้ของผู้ใช้แต่ละกลุ่มที่แตกต่างกันนั้นเป็นผลมาจากความยากง่ายในการใช้งานของผู้ใช้งาน และการเข้าถึงสารสนเทศทั้งที่เป็นรูปแบบของเว็บไซต์ เว็บแอปพลิเคชัน หรือโมบายแอปพลิเคชัน เป็นต้น จากข้อค้นพบดังกล่าวสามารถนำมากำหนดเป็นแนวทางการนำเสนอสารสนเทศด้วยภาพข้อมูลที่เหมาะสมตามหลักประสิทธิภาพของผู้ใช้งานในงานวิจัยชิ้นนี้ ดังภาพที่ 6 ได้แก่ 1) หลักการ Visual Design คือ การออกแบบหน้าตาของเว็บไซต์ให้สวยงาม เช่น การใช้เทคนิค Bootstrap เพื่อออกแบบเรสปอนซีฟเว็บไซต์ที่เหมาะสมกับการแสดงผลกับอุปกรณ์ที่มีขนาดหน้าจอแตกต่างกัน เป็นต้น 2) หลักการ Usability คือ การออกแบบที่เน้นความง่ายในการใช้งาน เช่น การใช้ปฏิทิน (Calendar Datepicker) สำหรับระบุข้อมูลในรูปแบบวันที่แทนการระบุข้อมูลแบบปกติ เพื่อลดโอกาสความผิดพลาดที่อาจจะเกิดขึ้นได้ เป็นต้น 3) หลักการ Interaction Design คือ รูปแบบการตอบสนองต่อการใช้งานเช่น การปฏิสัมพันธ์กับผู้ใช้เมื่อนำเมาส์ไปวางบนตามตำแหน่งของจังหวัดบนแผนที่ประเทศไทยแล้วแสดงข้อมูลการเกิดอุบัติเหตุ เป็นต้น และ 4) หลักการ Accessibility คือ การเพิ่มประสิทธิภาพในการเข้าถึง เช่น การใส่ความเข้มของสีจากสีขาว สีเหลือง สีส้ม และสีแดงในการแสดงผลตามปริมาณความถี่การเกิดอุบัติเหตุ เพื่อสื่อถึงระดับความรุนแรงของผลกระทบจากข้อมูล ตามแนวทาง Web Content Accessibility Guidelines (WCAG) [42] ของ W3C เป็นต้น

ข้อเสนอแนะ

ผลลัพธ์จากกระบวนการสกัดข้อมูลในงานวิจัยชิ้นนี้ ไม่สามารถตรวจสอบความถูกต้องด้วยการเปรียบเทียบกับสถิติการเกิดอุบัติเหตุจากแหล่งอ้างอิงใด ๆ ได้ สาเหตุมาจากข้อจำกัดด้านปริมาณของข้อมูลตั้งต้นจากแหล่งอ้างอิงที่ยังมีจำนวนไม่สอดคล้องกับความเป็นจริง เนื่องจากข้อมูลที่ปรากฏบนเว็บไซต์ของบริษัท

กลางคุ้มครองผู้ประสบภัยจากรถ จำกัด นั้นเก็บข้อมูลเฉพาะผู้ที่มาใช้สิทธิตามพระราชบัญญัติคุ้มครองผู้ประสบภัยจากรถเท่านั้น อย่างไรก็ตาม ผลลัพธ์ของงานวิจัยชิ้นนี้มุ่งนำเสนออีกแนวทางในการออกแบบและพัฒนาระบบสกัดข้อมูลกึ่งมีโครงสร้างที่มีความถูกต้องสูง สามารถนำไปประยุกต์ใช้สกัดข้อมูลลักษณะเดียวกันในอนาคต สิ่งที่ต้องพัฒนาต่อไปคือ การปรับปรุงคลังคำศัพท์ที่เกี่ยวข้องอยู่เสมอ เพื่อให้กระบวนการสกัดข้อมูลมีความถูกต้องยิ่งขึ้น โดยเฉพาะอย่างยิ่งสำหรับชุดข้อมูลภาษาไทย ตลอดจนนำออนโทโลยีเข้ามาช่วยอำนวยความสะดวกแก่ผู้ใช้ในการค้นคืนสารสนเทศด้วยหลักการค้นหาเชิงความหมาย ซึ่งจะดำเนินการต่อไปในอนาคต

เอกสารอ้างอิง

- [1] World Health Organization. (2018). *Global status report on road safety 2018*. Geneva: World Health Organization.
- [2] Division of Non Communicable Disease. (2015). *Where did the data and statistics of the dead disappear?*. Nonthaburi: Bureau of Non Communicable Disease.
- [3] Road Safety Center. (2017). *Accident data definition*. Bangkok: Department of Disaster Prevention and Mitigation, Ministry of Interior.
- [4] Hércules Antonio do Prado; and Edilson Ferneda. (2008). *Emerging Technologies of Text Mining: Techniques and Applications*. New York: Hershey.
- [5] Nitin Indurkha; and Fred J. Damerau. (2010). *Handbook of natural language processing*. 2nd ed. Cambridge: Chapman & Hall/CRC.
- [6] Fu, Guohong; and Luke, Kang-Kwong. (2005, June). Chinese named entity recognition using lexicalized HMMs. *ACM SIGKDD Explorations Newsletter*. 7(1): 19–25.
- [7] Tongtep, Nattapong; and Theeramunkong, Thanaruk. (2010, January-March). Pattern-based Extraction of Named Entities in Thai News Documents. *Thammasat International Journal of Science and Technology*. 15(1): (70-81).
- [8] Liu, Xiaohua; Wei, Furu; Zhang, Shaodian; and Zhou, Ming. (2013, February). Named entity recognition for tweets. *ACM Transactions on Intelligent Systems and Technology*. 4(1): 1-15.
- [9] Kanwal, Safia; Malik, Kamran; Shahzad, Khurram; Aslam, Faisal; and Nawaz, Zubair. (2019, June). Urdu Named Entity Recognition: Corpus Generation and Deep Learning Applications. *ACM Transactions on Asian and Low-Resource Language Information Processing*. 19(1): 1-13.
- [10] Wangtragulsang, Chinorot; Phaphoom, Nattakarn; Na Lamphun, Phannachet; Charnkeitkong, Pisit; and Qu, Jian. (2019, July-December). Thai Celebrity Information Extraction Based on Association Rule Measures. *International Scientific Journal of Engineering and Technology (ISJET)*, 3(2): 42-50.
- [11] Chantaraj, Pongkorn; and Rungrattanaubol, Jaratsri. (2563, January-June). Applied Information Extraction technique for extracting the king name who build a temple in Lanna Kingdom from historical documents. *Information Technology Journal*. 16(1): (24-33).
- [12] Chantrapornchai, Chantana; and Tunsakul, Aphisit. (2021, April). Information Extraction Tasks based on BERT and SpaCy on Tourism Domain. *ECTI Transactions on Computer and Information Technology*. 15(1): 108-122.
- [13] Ian Sommerville. (2010). *Software Engineering*. 9th ed. Boston: Addison-Wesley.

- [14] Alan Dennis; Barbara Haley Wixom; and Roberta M. Roth. (2012). *System Analysis and Design*. 5th ed. New York: John Wiley & Sons, Inc.
- [15] Phil Dutson. (2014). *Responsive Mobile Design Designing for Every Device*. Indiana: Addison-Wesley.
- [16] Ryan Mitchell. (2015). *Web Scraping with Python*. 2nd ed. Sebastopol: O'Reilly Media, Inc.
- [17] Simon Munzert; Christian Rubba; Peter Meißner; and Dominic Nyhuis. (2015). *Automated Data Collection with RA Practical Guide to Web Scraping and Text Mining*. New York: John Wiley & Sons, Inc.
- [18] Khalil, Salim; and Fakir, Mohamed. (2017, November). RCrawler: An R package for parallel web crawling and scraping. *SoftwareX*. 6: 98-106.
- [19] Office of Transport and Traffic Policy and Planning. (2562). *Ministry of Transport's road accident situation analysis report 2018*. Bangkok: Ministry of Transport.
- [20] Jorge Martinez. (2018). *Google Charts for Institutional Research Websites*. Houston: University of Houston.
- [21] Chun-houh Chen; Wolfgang Härdle; and Antony Unwin. (2008). *Handbook of Data Visualization*. Berlin: Springer-Verlag.
- [22] Claus O. Wilke. (2019). *Fundamentals of Data Visualization A Primer on Making Informative and Compelling Figures*. Sebastopol: O'Reilly Media, Inc.
- [23] The International Organization for Standardization (ISO). (1998). *Ergonomic requirements for office work with visual display terminals (VDTs) Part 11: guidance on usability*. Genève: ISO 1998.
- [24] Brian Shackel; and Simon Richardson. (1991). *Human factors for informatics usability – Background and overview*. Cambridge: Cambridge University Press.
- [25] Tom Brinck; Darren Gergle; and Scott D. Wood. (2012). *Usability for the web: Designing web sites that work*, San Francisco: Morgan Kaufmann.
- [26] Jennifer Preece. (2002). *Interaction design: Beyond human-computer interaction*, New York: John Wiley & Sons, Inc.
- [27] Alan Dix; Janet E. Finlay; Gregory D. Abowd; and Russell Beale. (2004). *Human-computer interaction*, 3rd ed. New Jersey: Pearson, Upper Saddle River.
- [28] Ben Shneiderman; and Catherine Plaisant. (2005). *Designing the user interface: Strategies for effective human-computer interaction*. 4th ed., Massachusetts: Addison Wesley Longman.
- [29] Jeff Rubin; and Dana Chisnell. (2008) *Handbook of usability testing: How to plan, design, and conduct effective tests*. Indiana: Wiley.
- [30] Jacob Nielsen. (1993) *Usability Engineering*. New York: Morgan Kaufmann.
- [31] Folmer, Eelke; and Bosch, Jan. (2004, February). Architecting for usability: a survey. *The Journal of Systems and Software*. 70(1-2): 61-78.
- [32] Rensis Likert. (1961). *New Patterns of Management*. New York: McGraw-Hill Book Company Inc.
- [33] W. Lawrence Neuman. (2012). *Basics of Social Research: Qualitative & Quantitative Approaches*. 3rd ed. Harlow: Pearson Education Limited.

- [34] Nielsen, Jakob; and Landauer, Tomas K. (1993). A Mathematical model of the finding of usability problems. In *the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*. pp. 206-213. Amsterdam: Association for Computing Machinery.
- [35] Boonchom Srisa - ard. (2554). *Basic Research*. Bangkok: Suveiriyasan Publishing.
- [36] Sukamolson, Suphat. (2560, July-December). Priori and Posteriori Comparisons for a Research Study. *Academic Journal of Buriram Rajabhat University*. 9(2): (51-70).
- [37] Wannaratn, Wirach. (2560, October 2016 - January 2017). Test Score and Grading. *Journal of Humanities and Social Sciences, Rajapruk University*. 2(3): (1-11)
- [38] Pasunon, Prasopchai. (2557, April - September). Reliability of Questionnaire in Quantitative Research. *Parichart Journal* 27(1): (145-163).
- [39] Santirattanaphakdi, Chakkarin; (2561, January-December). Online Marketing and Customer Service by Chatbot Case Study: Chatfuel in Customer Interactive on Messenger. *Sripatum review of science and technology*. 10: (71-87).
- [40] Chanwimalueng, Waiwit; and Polnigongit, Weerapong. (2561, January-June). A study of usability of elderly upon button size and shape on smartphone for creating fuzzy logic model. *Srinakharinwirot University (Journal of Science and Technology)*. 10(19): (121-135).
- [41] Rex Hartson and Pardha S. Pyla. (2012). *The UX Book Process and Guidelines for Ensuring a Quality User Experience*. Waltham: Elsevier, Inc.
- [42] Jeremy J. Sydik. (2007). *Design Accessible Web Sites Thirty-Six Keys to Creating Content for All Audiences and Platform*. Texas: The Pragmatic Bookshelf.