การวิเคราะห์สมการถดถอยพหุคูณแบบลำดับขั้นในการทดสอบความสามารถในการนำเสนอสารสนเทศการสกัดข้อมูลรายงานอุบัติเหตุทางถนน

จักรินทร์ สันติรัตนภักดี*
มหาวิทยาลัยวงษ์ชวลิตกุล ต.บ้านเกาะ อ.เมือง จ.นครราชสีมา 30000

** Corresponding Author: chakkarin_san@vu.ac.th*
อาจารย์ สาขาวิชาเทคโนโลยีธุรกิจดิจิทัล คณะบริหารธุรกิจ

ข้อมูลบทความ

ประวัติบทความ :
รับเพื่อพิจารณา : 12 พฤษภาคม 2564
แก้ไข : 2 มิถุนายน 2565
ตอบรับ : 22 สิงหาคม 2565
DOI : 10.14456/kmuttrd.2022.21

คำสำคัญ : ความสามารถในการใช้งาน / การสกัดข้อมูล / อุบัติเหตุทางถนน / การแสดงภาพข้อมูล

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสกัดข้อมูลรายงานอุบัติเหตุทางถนน และทดสอบความสามารถในการนำเสนอสารสนเทศที่เกี่ยวข้อง การดำเนินงานเริ่มจากการรวบรวมข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่จากเว็บไซต์ การเตรียมข้อมูลเพื่อการวิเคราะห์ข้อความ การจัดการกับคำที่เขียนไม่ถูกต้องด้วยการวัดความคล้ายคลึงของคำด้วยเทคนิคการวัดระยะทางเลเวนชเตย์น (Levenshtein Distance) เพื่อสร้างคลังข้อมูลคำศัพท์ร่วมกับการรู้จำเอนทิตี้มาสกัดข้อมูลใน 6 ประเด็น ได้แก่ จังหวัด วันที่ จำนวนผู้เสียชีวิต จำนวนผู้บาดเจ็บ ประเภทและจำนวนของยานพาหนะที่เกิดเหตุ เพื่อนำเสนอสารสนเทศด้วยภาพข้อมูลปฏิสัมพันธ์กับผู้ใช้ผ่านทางเว็บไซต์ เมื่อทดสอบความสามารถในการนำเสนอสารสนเทศรายงานอุบัติเหตุทางถนนในแง่ของประสิทธิผลการค้นคืนข้อมูล พบว่า ค่าความถูกต้อง ค่าความครบถ้วน และค่าประสิทธิภาพโดยรวม เท่ากับ 0.87, 0.85 และ 0.86 ตามลำดับ ผลการทดสอบความสามารถในการนำเสนอสารสนเทศจากผู้ใช้ 30 คน จากการสุ่มตัวอย่างโดยไม่ใช้ความน่าจะเป็นด้วยวิธีการเลือกกลุ่มตัวอย่างแบบเฉพาะเจาะจงอยู่ในระดับดี ($\bar{X}$ = 3.73) ก่อนการวิเคราะห์ข้อมูล ตรวจสอบข้อตกลงเบื้องต้นในการวิเคราะห์การถดถอย เมื่อวิเคราะห์การถดถอยพหุคูณด้วยวิธีคัดเลือกแบบลำดับขั้น พบว่า ประกอบด้วย 6 ตัวแปร ได้แก่ ความถูกต้องของผลการสกัดข้อมูลรายงานอุบัติเหตุ ความสามารถในการใช้งานระบบได้อย่างรวดเร็ว ความถูกต้องของการแสดงผลข้อมูลตามเงื่อนไขการค้นหา ความสามารถในการทำความเข้าใจกับกระบวนการได้มาซึ่งผลลัพธ์ได้ด้วยตนเอง ความถูกต้องของการจัดวางรายชื่อจังหวัดเพื่อการค้นหา และความรวดเร็วในการนำเสนอภาพข้อมูล ตัวแปรทั้งหมดมีความแม่นยำในการพยากรณ์ผลการทดสอบความสามารถในการนำเสนอสารสนเทศโดยรวมเท่ากับ 87.2 และมีค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ .934 ซึ่งมีนัยสำคัญทางสถิติ องค์ประกอบการใช้งานในส่วนที่เป็นสมรรถนะของผู้ใช้มีแนวโน้มไปในทิศทางเดียวกัน ตรงกันข้ามกับส่วนที่เป็นความคิดเห็นของผู้ใช้ ซึ่งแตกต่างกันไปตามบริบท

Stepwise Multiple Regression Analysis for Usability Testing of Presented Information on Extracted Road Traffic Accident Reports

Chakkarin Santirattanaphakdi*
Vongchavalitkul University, Banko, Mueang, Nakhon Ratchasima 30000

* *Corresponding Author: chakkarin_san@vu.ac.th*
Lecturer, Department of Digital Business Technology, Faculty of Business Administrator.

Article Info

Article History:
Received: May 12, 2021
Revised: June 2, 2022
Accepted: August 22, 2022
DOI : 10.14456/kmuttrd.2022.21

Keywords : Usability Testing / Data Extraction / Road Traffic Accident / Data Visualization

Abstract

This research aimed to extract road traffic accident reports and perform usability testing of the presented relevant information. The process started from collecting data on large-scale accidents from websites, pre-processing the data for text analysis and handling typographical errors by measuring text similarity using the Levenshtein distance for creating a specialized corpus and entity extraction to extract relevant information in 6 issues, namely, province, date, number of deaths, number of injuries, type and number of motor vehicle(s) involved in the accidents. The information was then visually presented to the web user via data visualization. When evaluating the information retrieval by three indicators, i.e., Precision, Recall, and F-measure, the average accuracy of information retrieval for the tested samples were 0.87, 0.85, and 0.86, respectively. The result on usability testing of the presented information, based on 30 users selected by non-probability sampling via the purposive sampling method, was at a good level ($\bar{X}$ = 3.73). Prior to the analysis of the data, the assumptions of linear regression were checked. When analyzing multiple regression with stepwise selection procedures, 6 variables were noted to be involved: the accuracy of the extracted road traffic accident reports, the ability to rapidly use the system, the accuracy of the presented results as per the search criteria, the ability of understand the process of obtaining results on one's own, the accuracy of the placement of the list of provinces for searching and the speed at which the data are presented. All variables exhibited the prediction accuracy of the overall information presentation capability of 87.2, with multiple coefficients value of .934, which was significant. The user performance components tended to be in the same direction, while the user opinions components varied with the relevant context.

1. บทนำ

จากการสำรวจขององค์การอนามัยโลก พบว่า ในปี 2018 ประเทศไทยมีอัตราการเกิดอุบัติเหตุทางถนนสูงเป็นอันดับ 9 ของโลก มีคนไทยเสียชีวิตจากอุบัติเหตุทางถนนปีละประมาณ 22,491 ราย คิดเป็น 32.7 คนต่อประชากร 1 แสนคน เฉลี่ยแล้วมีคนไทยเสียชีวิตจากอุบัติเหตุชั่วโมงละ 3 คน [1] ซึ่งถือว่าเกินค่าเฉลี่ยของโลกถึง 2 เท่า อุบัติเหตุทางถนนไม่เพียงส่งผลกระทบต่อผู้ที่ประสบภัยเท่านั้น แต่ยังก่อให้เกิดความสูญเสียต่อเศรษฐกิจและสังคมไทยโดยรวมอย่างมหาศาล

ประเทศไทยมีการเก็บข้อมูลเชิงสถิติเพื่อนำมาวิเคราะห์ถึงแนวโน้มสถานการณ์อุบัติเหตุทางถนน รวมถึงกำหนดทิศทางนโยบายต่าง ๆ ทั้งระยะเร่งด่วนและระยะยาว แต่ปัจจุบันข้อมูลที่นำมาใช้มาจากหลายหน่วยงาน และมีวัตถุประสงค์ในการเก็บข้อมูลแตกต่างกัน [2] ทำให้ข้อมูลจำนวนผู้บาดเจ็บและเสียชีวิตคลาดเคลื่อน และไม่มีข้อมูลจากหน่วยงานใดที่มีความครอบคลุมครบถ้วน จากการศึกษาระบบจัดเก็บข้อมูลการเกิดอุบัติเหตุร่วมกับการสังเคราะห์เอกสาร และงานวิจัยที่เกี่ยวข้อง พบว่า ส่วนใหญ่เน้นจัดเก็บเพียงข้อมูลเชิงปริมาณเท่านั้น แต่ยังมีข้อมูลเชิงคุณภาพที่น่าเชื่อถือและยังรอการนำไปใช้ประโยชน์อีกรูปแบบหนึ่ง นั่นคือการรายงานอุบัติเหตุรายใหญ่ (Major Accident) [3] ที่มีผู้เสียชีวิต 2 รายขึ้นไป หรือผู้บาดเจ็บ 4 รายขึ้นไป หรือผู้บาดเจ็บและผู้เสียชีวิตรวม 4 ราย ขึ้นไป ตลอดจนอุบัติเหตุทางถนนที่เป็นที่สนใจของประชาชน เช่น บุคคลสำคัญหรือบุคคลสาธารณะเกิดบาดเจ็บ พักรักษาตัวที่โรงพยาบาล หรือเสียชีวิต หรือรถน้ำมัน รถสารเคมีเกิดอุบัติเหตุ จนเกิดเพลิงไหม้ หรือสารเคมีฟุ้งกระจายไปสู่ชุมชน เป็นต้น ที่รายงานมายังศูนย์รับแจ้งอุบัติเหตุและประสานการช่วยเหลือผู้ประสบภัยจากรถ บริษัทกลางคุ้มครองผู้ประสบภัยจากรถจำกัด เผยแพร่ต่อสาธารณะผ่านเว็บไซต์ www.thairsc.com เป็นข้อมูลเพื่อนำไปเยียวยาผู้ประสบภัยจากรถให้ได้รับการคุ้มครองตามกฎหมาย และเป็นแนวทางในการป้องกันและลดการเกิดอุบัติเหตุทางถนนในรูปแบบของข้อมูลประเภทกึ่งมีโครงสร้าง จากการกำหนดหัวข้อในการบันทึกข้อมูลโดยแต่ละช่องเปิดโอกาสให้ผู้รายงานสามารถใส่ข้อมูลตามสถานการณ์ที่เกิดขึ้นจริง ทำให้ในแต่ละครั้งข้อมูลในการรายงานขาดความถูกต้อง ไม่เป็นมาตรฐาน และแตกต่างกันไปตามบริบทของผู้รายงาน เนื่องจากผู้ใช้ต้องวิเคราะห์ และสังเคราะห์ข้อมูลด้วยตนเองก่อนจึงจะสามารถนำไปใช้ประโยชน์ จึงเผยแพร่แก่ผู้ใช้เพียงในวงจำกัด ผู้วิจัยทราบถึงปัญหา และเล็งเห็นถึงความสำคัญของข้อมูล จึงออกแบบและพัฒนาการสกัดข้อมูลรายงานอุบัติเหตุรายใหญ่บนเว็บไซต์เพื่อนำเสนอสารสนเทศแก่ผู้ใช้ด้วยเทคนิคภาพข้อมูล

อย่างไรก็ดี บุคคลย่อมมีความแตกต่างในการรับสารสนเทศ [4] ตามความต้องการใช้สารสนเทศเกิดขึ้น และเปลี่ยนแปลงอยู่ตลอดเวลา และความแตกต่างนั้นมีผลสืบเนื่องมาจากปัจจัยที่เกี่ยวกับองค์การและสภาพแวดล้อม ปัจจัยที่เกี่ยวกับตัวบุคคลหรือผู้ใช้ และปัจจัยที่เกี่ยวกับลักษณะของสารสนเทศ ไม่ว่าจะเป็นความแตกต่างในการรับรู้รูปร่าง ขนาด สี หรือเนื้อหาที่สัมพันธ์กับคุณลักษณะทางกายภาพของผู้รับสาร เช่น ประสิทธิภาพการรับสารสนเทศในลักษณะรูปภาพ หรือข้อความขึ้นอยู่กับความเสื่อมของดวงตา [5] เป็นต้น ตลอดจนความแตกต่างในการรับสารสนเทศ อันเกิดจากคุณลักษณะด้านประชากรศาสตร์ เช่น อายุ การศึกษา ตลอดจนประสบการณ์ ล้วนส่งผลจากปัจจัยที่เกี่ยวกับองค์การและสภาพแวดล้อมทั้งสิ้น แต่ขณะที่งานวิจัยจำนวนมากมุ่งนำเสนอสารสนเทศด้วยเทคนิคภาพข้อมูล และประเมินคุณภาพของสารสนเทศที่นำเสนอด้วยการวัดความพึงพอใจเท่านั้น ซึ่งเป็นการวัดผลเพียงมิติเดียว ในทางกลับกันคุณภาพของสารสนเทศมีหลายองค์ประกอบที่บูรณาการกัน งานวิจัยนี้จึงนำการประเมินความสามารถในการใช้งานเข้ามาช่วยวัดคุณภาพของการนำเสนอสารสนเทศผ่านเว็บไซต์ และวิเคราะห์ความสัมพันธ์แต่ละองค์ประกอบด้วยสมการถดถอยพหุคูณแบบลำดับขั้น อันจะเป็นแนวทางในการนำเสนอสารสนเทศที่เหมาะสมต่อไปในอนาคต

2. วัตถุประสงค์

2.1 เพื่อสกัดข้อมูลรายงานอุบัติเหตุทางถนน

2.2 เพื่อทดสอบความสามารถในการนำเสนอสารสนเทศการสกัดข้อมูลรายงานอุบัติเหตุทางถนน

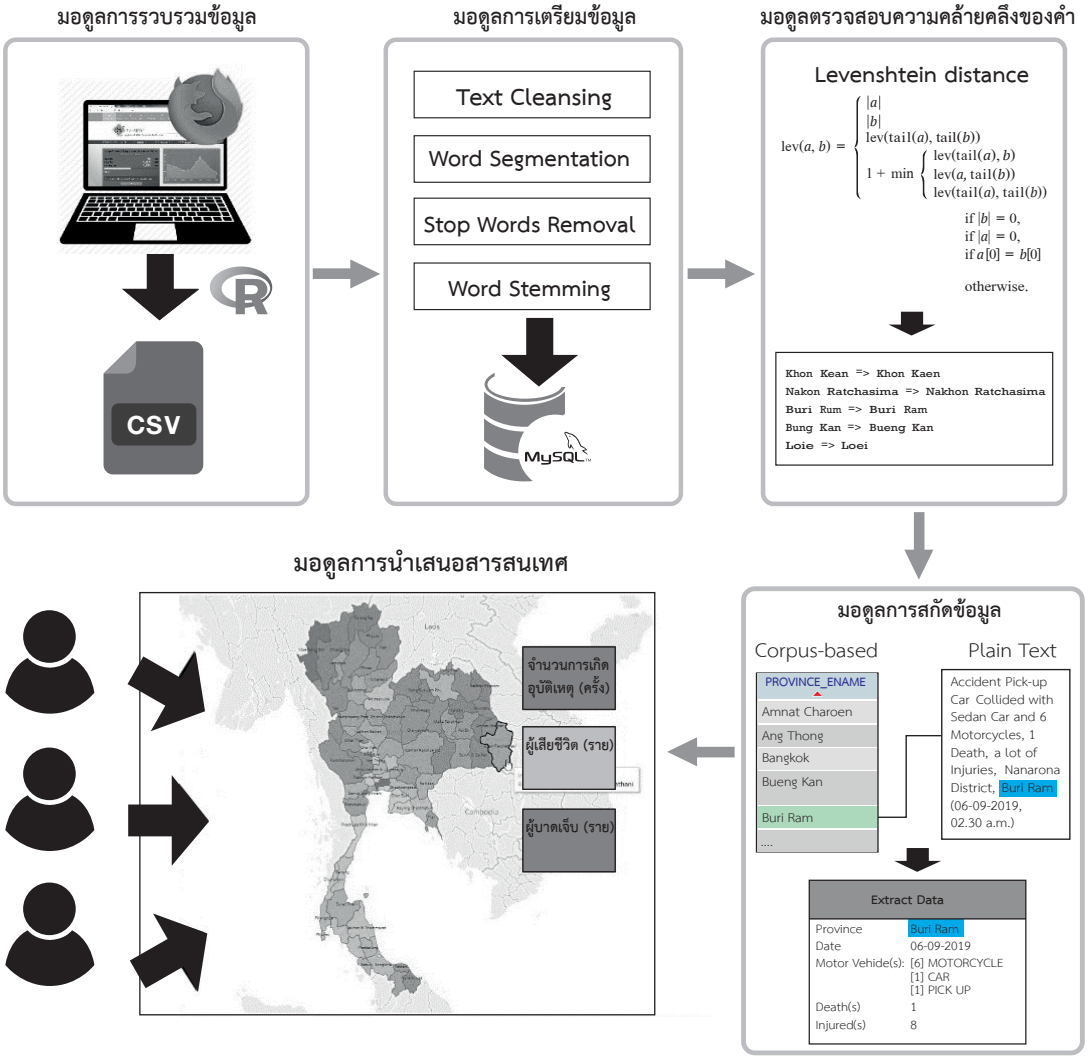
3. สมมติฐานการวิจัย

3.1 ผลการทดสอบความสามารถในการนำเสนอสารสนเทศรายงานอุบัติเหตุทางถนนด้านประสิทธิผลของการค้นคืนข้อมูล มีค่าความถูกต้องและค่าความครบถ้วนมากกว่าร้อยละ 80

3.2 ผลการทดสอบความสามารถในการนำเสนอสารสนเทศการสกัดข้อมูลรายงานอุบัติเหตุทางถนน มากกว่า 3.51

4. วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นงานวิจัยเชิงประยุกต์ โดยใช้แนวคิดแบบจำลองการพัฒนาแอปพลิเคชันแบบเร่งรัด (Rapid Application Development Model: RAD) เพื่อช่วยให้การพัฒนาระบบสามารถดำเนินการได้อย่างรวดเร็ว [6] โดยแยกการพัฒนาออกเป็นมอดูล ก่อนจะนำมาประกอบเป็นชิ้นงานที่สมบูรณ์ ประกอบด้วย 5 มอดูล ได้แก่ มอดูลการรวบรวมข้อมูล มอดูลการเตรียมข้อมูล มอดูลการตรวจสอบความคล้ายคลึงของคำ มอดูลการสกัดข้อมูล และมอดูลการนำเสนอสารสนเทศ ดังรูปที่ 1 มีรายละเอียดดังนี้



รูปที่ 1 กรอบการดำเนินงานการสกัดข้อมูลรายงานอุบัติเหตุทางถนน

4.1 มอดูลการรวบรวมข้อมูล (Data Collection Module) จากเว็บไซต์ ด้วยเทคนิคการขูดเว็บ (Web Scraping) ที่มีหลักการทำงานเช่นเดียวกับการเข้าใช้งานเว็บไซต์ผ่านโปรโตคอลเอชทีทีพี (Hypertext Transfer Protocol: HTTP) [7] ดังนั้นการขูดเว็บจะทำงานที่เสมือนกับว่าเข้าใช้เว็บไซต์ด้วยตนเอง แต่ในทางกลับกันจะเขียนคำสั่งหรือใช้โปรแกรมสำเร็จรูปในการเข้าถึงเว็บไซต์ และคัดลอกข้อมูลในตำแหน่งที่ระบุมาจัดเก็บในรูปของไฟล์ที่รอการนำไปประมวลผล

อย่างไรก็ดี มีข้อควรระวัง 2 ประการ คือ มิให้สร้างภาระในการทำงานของเครื่องแม่ข่ายจนอาจก่อให้เกิดความเสียหาย และการอนุญาตให้เข้าถึงและคัดลอกข้อมูลบนเว็บไซต์ ดังนั้นผู้วิจัยทำการทดสอบว่าเว็บไซต์ www.thairsc.com นั้นอนุญาตการใช้งานของเว็บครอว์เลอร์ (Web Crawler) ซึ่งเป็นโปรแกรมอัตโนมัติที่ใช้ในการเก็บข้อมูลจากเว็บไซต์ด้วยการเขียนโปรแกรมภาษา R และใช้แพ็คเกจ robots.txt ในการทดสอบ ปรากฏผลลัพธ์คือการคืนค่า TRUE แสดงถึงเว็บไซต์ดังกล่าวสามารถใช้เทคนิคเว็บครอว์เลอร์ในการเก็บข้อมูลได้ จึงทำการเขียนโปรแกรมภาษา R อีกครั้งด้วยแพ็คเกจ Rvest และ xml2 เพื่อเข้าถึงเว็บเพจแต่ละหน้าตามที่กำหนดไว้แบบวนลูป จากนั้นจะคัดลอกข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่ที่อยู่ในรูปแบบภาษาอังกฤษที่ปรากฏบนเว็บไซต์ www.thairsc.com ตั้งแต่วันที่ 1 มกราคม 2559 - 31 มีนาคม 2563 รวมทั้งสิ้น 3,025 รายการ แล้วนำมาบันทึกเป็นไฟล์นามสกุล .CSV (Comma Separated Value) แบ่งเป็นส่วนหัว (Header) และส่วนรายละเอียด (Description)

4.2 มอดูลการเตรียมข้อมูล (Data Preparation Module) เบื้องต้นผู้วิจัยทำการคัดกรองข้อมูล เพื่อให้ได้ข้อมูลที่มีความสมบูรณ์ โดยตรวจสอบโครงสร้างของข้อมูลจากขั้นตอนที่ผ่านมา พบว่า รายงานอุบัติเหตุรายใหญ่ที่อยู่ในรูปแบบภาษาอังกฤษจำนวนหนึ่งยังขาดความสมบูรณ์ในด้านเนื้อหาเมื่อเทียบกับรายงานรูปแบบภาษาไทย ดังนั้นจึงพิจารณาตัดข้อมูลที่ไม่ตรงตามเงื่อนไขออก ตามหลัก Listwise deletion [8] คงเหลือข้อมูลทั้งสิ้น 1,353 รายการ จากนั้นเตรียมข้อมูลตามกระบวนการ ดังนี้

4.2.1 การทำความสะอาดข้อความ (Text Cleaning) โดยจัดการกับเครื่องหมาย หรืออักขระพิเศษที่ใช้ในการแสดงผล ค้นหาและแทนที่ตัวอักษรภาษาไทยให้เป็นภาษาอังกฤษทั้งหมด ปรับรูปแบบของวันที่ให้อยู่ในรูปแบบมาตรฐาน ตลอดจนตัดข้อความส่วนท้ายของทุก ๆ รายการออก เนื่องจากเป็นเพียงรายละเอียดที่ตั้งและช่องทางการติดต่อ

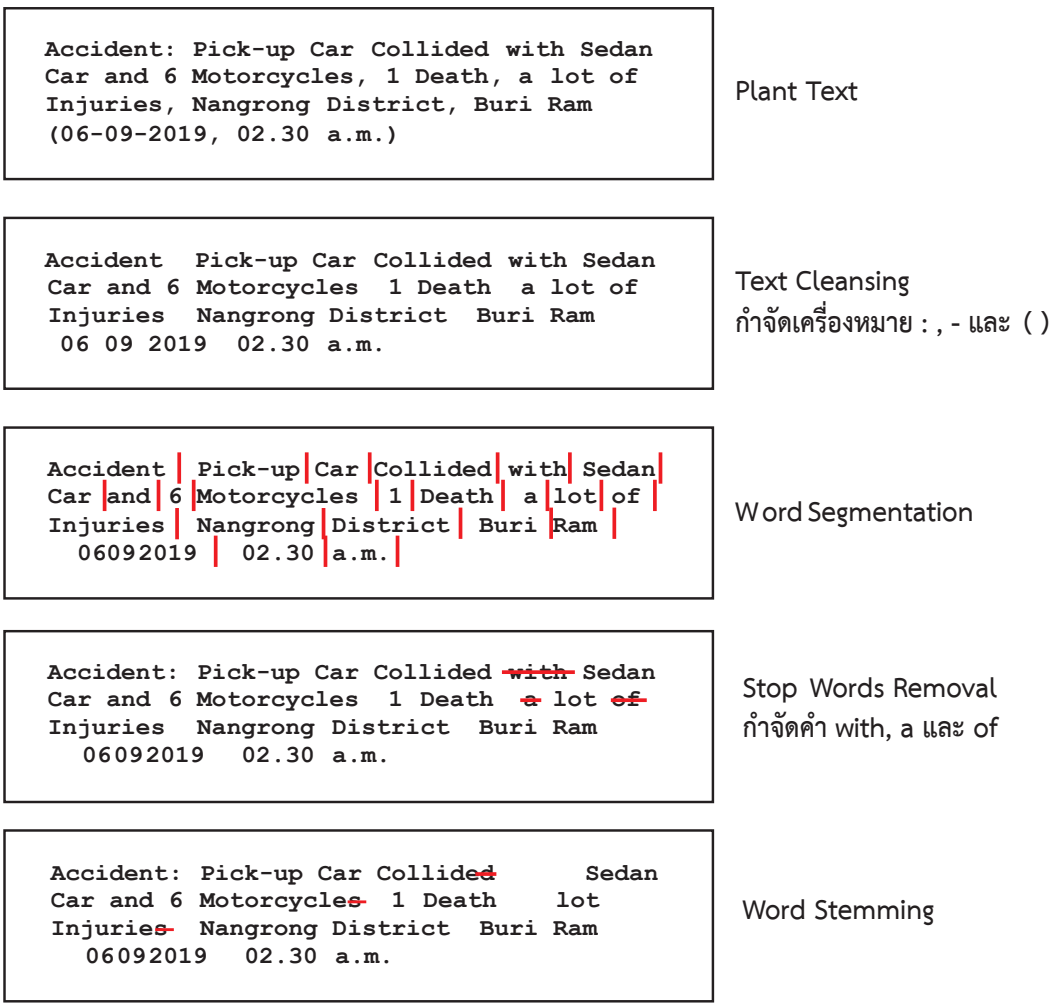
4.2.2 การตัดคำ/ข้อความ (Word Segmentation) เป็นการนำข้อความทั่วไปซึ่งอยู่ในรูปแบบประโยคมาแบ่งออกเป็นคำหรือคุณลักษณะ (Term/Feature) งานวิจัยในปัจจุบันได้เสนอวิธีการตัดคำ แบ่งตามกระบวนการทำงานเป็น 3 กลุ่ม [9] ได้แก่ 1) การตัดคำโดยใช้กฎ (Rule-Based Approach) เป็นการตัดคำโดยใช้วิธีเกณฑ์ทางอักขรวิธีที่กำหนดลักษณะของการประสมอักษร 2) การตัดคำโดยใช้พจนานุกรม (Dictionary-Based Approach) ที่ป้อนข้อความเข้าไปค้นหา และเปรียบเทียบสายอักขระกับคำศัพท์ในพจนานุกรม เพื่อหาว่าข้อความดังกล่าวควรตัดคำในบริเวณใด และประกอบด้วยคำใดบ้าง เช่น พจนานุกรมคำศัพท์อิเล็กทรอนิกส์ (LEXiTRON) อย่างไรก็ดี การตัดคำโดยใช้พจนานุกรมก็มีข้อจำกัดบางประการ เนื่องจากมีความเป็นไปได้ที่คำที่ปรากฏในเอกสาร อาจจะไม่ปรากฏในพจนานุกรม จึงเป็นที่มาของเอ็นแกรม (N-gram) ที่นำบางส่วนของข้อความออกมาเป็นตามค่า N และ 3) การตัดคำโดยใช้คลังข้อมูล (Corpus-Based Approach) โดยเตรียมคลังข้อมูลที่มีการตัดคำและการกำกับหน้าที่ของคำไว้ล่วงหน้าซึ่งในประโยคภาษาอังกฤษใช้เว้นวรรคในการตัดคำ จึงไม่ประสบปัญหาของการแยกคำแบบภาษาไทย

4.2.3 การกำจัดคำหยุด (Stop Word Removal) เป็นการกำจัดคำที่ไม่มีนัยสำคัญออกไป โดยไม่ทำให้ความหมายในเอกสารหรือข้อความนั้นเสียหรือเปลี่ยนแปลงไป โดยมากจะเป็นคำนำหน้านาม (Article) คำบุพบท (Preposition) คำสรรพนาม (Pronoun) และคำสันธาน (Conjunction) เช่น “a”, “an”, “the”, “he”, “she”, “it”, “and”, “who”, “which”, และ “that” เป็นต้น

4.2.4 การหารากศัพท์ของคำ (Word Stemming) เป็นการหารูปแบบดั้งเดิมหรือรากศัพท์ของคำนั้น ๆ โดยตัดส่วนหน้า (Prefix) หรือส่วนท้าย (Suffix) ของคำออก

ซึ่งคำศัพท์ที่มีรากศัพท์เดียวกันจะมีลักษณะที่คล้ายคลึงกัน ส่วนใหญ่จะเป็นคำนามและคำกริยา เช่น คำว่า “crashed” รูปแบบเดิม แม้ว่า “ed” จะถูกตัดออกไป แต่ความหมายของคำยังคงเดิม การหารากศัพท์จะช่วยรวมคำดังกล่าวให้เป็นคำเดียวกัน เพื่อลดความซ้ำซ้อนของคำหรือคุณลักษณะที่เกิดขึ้นในเอกสาร ซึ่งงานวิจัยนี้ได้ใช้ขั้นตอนวิธีการหารากศัพท์ของ Porters Stemmer [10] เป็นวิธีการใช้กระบวนการ 5 ขั้นตอน และกฎจำนวน 60 กฎ เพื่อค้นหารากศัพท์ของคำ จุดเด่นคือมีผลลัพธ์การค้นหารากศัพท์ของคำที่มีความถูกต้องสูง มีข้อผิดพลาดน้อย ซึ่งเหมาะสำหรับการหารากศัพท์ของคำที่เป็นภาษาอังกฤษ

เมื่อดำเนินการเตรียมข้อมูลเสร็จสิ้นจะเหลือเฉพาะข้อความทั่วไป (Plain Text) ที่พร้อมนำไปประมวลผล ดังรูปที่ 2 กระบวนการดังกล่าวจะช่วยลดระยะเวลา และลดการใช้ทรัพยากรระบบในการประมวลผลอีกด้วย จากนั้นจะทำการนำเข้า (Import) ข้อมูลจำนวน 1,353 รายการ สู่ฐานข้อมูล MySQL เพื่อนำไปประมวลผลในขั้นตอนต่อไป



รูปที่ 2 ข้อความทั่วไปที่ผ่านกระบวนการเตรียมข้อมูล

4.3 มอดูลการตรวจสอบความคล้ายคลึงของคำ (Word Similarity Module) เนื่องจากเป็นการแจ้งข้อมูลรายงานอุบัติเหตุจากผู้รับผิดชอบ จึงมักเกิดปัญหาความหลากหลายของข้อความที่แตกต่างกันตามบริบทของผู้ใช้งาน ตลอดจนเกิดความผิดพลาดในการใช้ภาษาที่เกิดจากความตั้งใจหรือแม้แต่ไม่ตั้งใจของผู้ใช้ ส่งผลให้ผลลัพธ์อาจเกิดความผิดพลาดในการนำไปสร้างคลังคำศัพท์เพื่อสกัดข้อมูล โดยเฉพาะอย่างยิ่งรายชื่อจังหวัดที่มักพบความผิดพลาดเมื่ออยู่ในรูปแบบภาษาอังกฤษ หรือคำที่เขียนไม่ครบถ้วน ตลอดจนคำที่ไม่เป็นไปตามมาตรฐาน ดังตารางที่ 1

ตารางที่ 1 ตัวอย่างรายชื่อจังหวัดที่มักพบความผิดพลาด

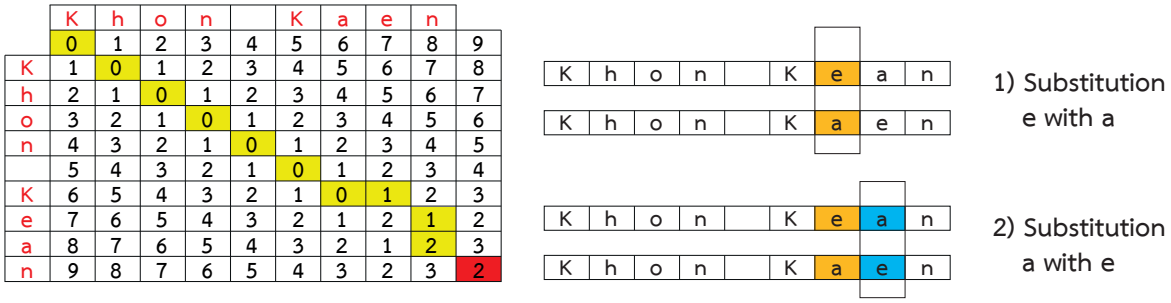
คำที่เขียนผิดพลาด	คำที่เขียนไม่ครบถ้วน	คำที่ไม่เป็นไปตามมาตรฐาน	คำที่ถูกต้อง
Khon Kean			Khon Kaen
Nakon Ratchasima		Nakhonratchasima	Nakhon Ratchasima
Buri Rum		Buri Ram	Buriram
Bung Kan			Bueng Kan
	Ayutthaya		Phra Nakhon Si Ayutthaya
....	....	....	....

ดังนั้นจะเห็นได้ว่าความถูกต้องของคำศัพท์จะส่งผลอย่างยิ่งต่อความถูกต้องของการสกัดข้อมูล [11] อย่างไรก็ดีหากเก็บคำศัพท์ที่ผิดพลาดเข้าสู่ฐานข้อมูลร่วมกับคำศัพท์ที่ถูกต้อง นอกจากจะทำให้คลังคำศัพท์มีขนาดใหญ่เกินไปเนื่องจากความซ้ำซ้อนของคำ และยากที่จะระบุได้ว่าจำนวนเท่าไรจึงจะครอบคลุมและครบถ้วน อีกทั้งยังปริมาณของข้อมูลจะส่งผลต่อการใช้ระยะเวลา และทรัพยากรระบบในการประมวลผล งานวิจัยนี้กำหนดรายชื่อจังหวัดในประเทศไทยในรูปแบบภาษาอังกฤษที่รับรองโดยราชบัณฑิตยสถาน เพื่อเปรียบเทียบความคล้ายคลึงกันของอักขระ (String Similarity) ด้วยการหาค่าความต่างกันของสายอักขระสองชุดด้วยเทคนิคการวัดระยะทางเลเวนชเตย์น (Levenshtein Distance) [12] ที่แก้ปัญหาโดยใช้กำหนดการพลวัต โดยค่าความต่างกันจะวัดจากจำนวนของอักขระที่จะต้องทำการแทรกเพิ่ม (Insertion) การลบ (Deletion) และการแทนที่ (Substitute) ของตัวอักษรหนึ่งตัว ระยะทางจะคำนวณจากจำนวนน้อยที่สุดของระยะทางที่แก้ไข (Edit Operation) [13] เพื่อแสดงความคล้ายคลึงที่มากที่สุด

งานวิจัยนี้ใช้การเขียนโปรแกรมภาษา PHP โดยนำผลลัพธ์จากการตัดข้อความมาเปรียบเทียบกับคำศัพท์รายชื่อจังหวัดที่กำหนดไว้ในฐานข้อมูล เพื่อวิเคราะห์ความคล้ายคลึงของคำศัพท์แต่ละคำ แล้วเลือกแสดงคำศัพท์ที่มีระยะทางแก้ไขน้อยที่สุดออกมา ดังรูปที่ 3

กระบวนการเปรียบเทียบความคล้ายคลึงกันของอักขระด้วยการหาค่าความต่างกันของสายอักขระสองชุดด้วยเทคนิคการวัดระยะทางเลเวนชเตย์น โดยนำผลลัพธ์จากการตัดข้อความมาเปรียบเทียบกับคำศัพท์ที่ดึงมาจากฐานข้อมูลและเก็บไว้ในรูปแบบอาร์เรย์ แล้วทำการเปรียบเทียบทีละคู่แบบวนลูป จากรูปที่ 3 เป็นการเปรียบเทียบระหว่างคำที่มาจากตัดข้อมูลคือ “Khon Kean” กับคำว่า “Khon Kaen” ที่กำหนดไว้ในฐานข้อมูล ซึ่งทั้งสองมีจำนวนความยาวอักขระเท่ากันคือ 9 ตัวอักษร กระบวนการเริ่มจากแยกอักษรแต่ละตัวออกจากกัน เพื่อเปรียบเทียบตัวอักษรทีละคู่ พบว่าตัวอักษรตัวที่ 1 - 6 ของทั้งสองข้อความตรงกัน ในขณะที่อักษรตัวที่ 7 ของผลลัพธ์จากการตัดข้อความคือ “e” เมื่อเปรียบเทียบกับคำที่กำหนดไว้ในฐานข้อมูลคือ “a” จึงต้องทำการแทนที่ตัวอักษร “e” ด้วย “a” และตรวจสอบตัวอักษรคู่ต่อไปด้วยกระบวนการเดียวกันในอักษรตัวที่ 8 ของผลลัพธ์จากการตัดข้อความคือ “a” เมื่อเปรียบเทียบกับคำที่กำหนดไว้ในฐานข้อมูลคือ “e” จึงต้องทำการแทนที่ตัวอักษร “a” ด้วย “e” ตามที่กำหนดไว้ จากนั้นจะทำการเปรียบเทียบจนครบทุกตัวอักษร หากแต่การเปรียบเทียบยังไม่จบเพียงเท่านี้ การนำผลลัพธ์จากการตัดข้อความ “Khon Kean” ที่เป็นคำที่ผิดพลาดจะทำการเปรียบเทียบกับชุดอักขระทั้งหมด

ที่กำหนดไว้ ในตัวอย่างคือรายชื่อจังหวัดจำนวน 77 จังหวัด จากนั้นจะตรวจสอบดูว่าผลการเปรียบเทียบชุดอักขระคู่ใดให้จำนวนระยะทางที่แก้ไขน้อยที่สุด ดังนั้นคำว่า “Khon Kean” จึงมีค่าความคล้ายคลึงกับคำว่า “Khon Kaen” ที่กำหนดไว้ในฐานข้อมูลเมื่อเทียบกับรายชื่อจังหวัดอื่น ๆ เนื่องจากมีระยะทางแก้ไขน้อยที่สุดเท่ากับ 2 ตามผลลัพธ์ของเมตริกซ์ดังรูปที่ 3



รูปที่ 3 การหาค่าความต่างกันของสายอักขระสองชุดด้วยเทคนิคการวัดระยะทางเลเวนชเตย์น

4.4 มอดูลการสกัดข้อมูล (Data Extraction Module) เริ่มจากการสร้างคลังข้อมูลคำศัพท์รายชื่อจังหวัดไประเทศไทยในรูปแบบภาษาอังกฤษ อันจะนำมาเป็นเงื่อนไขในการสกัดข้อมูล จากนั้นนำประเภทของยานพาหนะตามการจำแนกประเภทจากรายงานการวิเคราะห์สถานการณ์อุบัติเหตุทางถนนของกระทรวงคมนาคม ได้แก่ รถจักรยานยนต์ (Motorcycle) รถยนต์นั่ง (Sedan Car) รถโดยสารขนาดเล็ก (Pick-up Car) รถบรรทุก 10 ล้อหรือมากกว่า (Truck) คนเดินเท้า (Pedestrian) รถจักรยาน (Bicycle) รถแท็กซี่ (Taxi) รถบรรทุก 6 ล้อ (Mini Truck) รถโดยสารขนาดเล็ก (Van) และรถโดยสารขนาดใหญ่ (Bus) มาสร้างคลังคำศัพท์และกำหนดน้ำหนักของคำศัพท์ตามจำนวนตัวอักษร เพื่อเป็นเงื่อนไขการเรียงลำดับในการประมวลผล

จากนั้นกำหนดคุณลักษณะข้อความด้วยเรกูลาร์เอ็กซ์เพรชชัน (Regular Expression) เพื่อกำหนดรูปแบบหรือกลุ่มคำ (Pattern) ที่กำหนดขึ้นเพื่อค้นหาข้อความหรือตัวอักษรต่าง ๆ หรือสำหรับตรวจสอบว่าตรงตามเงื่อนไข (Condition) ที่กำหนดไว้หรือไม่ ในงานวิจัยนี้กำหนด 6 คุณลักษณะ ได้แก่ จังหวัด (Province) วันที่ (Date of Accident) จำนวนผู้เสียชีวิต (List of Deaths) จำนวนผู้บาดเจ็บ (List of Injured) ประเภท และจำนวนของยานพาหนะที่เกิดเหตุ (Number and Motor Vehicle(s) involved in the accident) โดยจะไม่นำข้อมูลส่วนบุคคลของผู้ประสบเหตุมาเป็นส่วนหนึ่งของผลลัพธ์จากการสกัดข้อมูล โดยสร้างเงื่อนไขในการค้นหารายชื่อจังหวัดจากส่วนหัวของรายงานอุบัติเหตุแต่ละรายการ หากไม่พบข้อมูลจะไปค้นหาในส่วนของรายละเอียด โดยยึดตามลำดับแรกที่พบในเขตข้อมูล และสร้างเงื่อนไขในการค้นหาประเภทของยานพาหนะที่เกิดเหตุจากส่วนของรายละเอียดแบบวนลูป เริ่มจากคำศัพท์ที่มีความยาวมากที่สุดไปจนถึงน้อยที่สุด ตามหลักการรู้จำชื่อเอนทิตี้ (Named Entity Recognition: NER) ที่เป็นหนึ่งวิธีการสกัดสารสนเทศที่ได้รับความนิยมอย่างมากในปัจจุบัน เนื่องจากมีกระบวนการทำงานที่ไม่ซับซ้อน และได้ผลลัพธ์ที่มีประสิทธิภาพ จึงถูกนำไปประยุกต์ใช้กับการวิเคราะห์ข้อความได้หลากหลาย โดยเฉพาะกับข้อมูลแบบเฉพาะเจาะจง เช่น ชื่อบุคคล ชื่อจังหวัด ชื่อประเทศ เป็นต้น [14] โดยใช้การทำป้ายลำดับ (Sequence Labeling) ในลักษณะของการหา และระบุ ตลอดจนแยกประเภทและส่วนประกอบของข้อความให้อยู่ในประเภทที่ต้องการตามกฎ (Rule-Based) แบ่งเป็น 2 รูปแบบ คือ 1) ค้นหาและระบุตำแหน่งข้อความตามเงื่อนไขที่ระบุไว้ล่วงหน้าในคลังคำศัพท์ที่สร้างขึ้น เพื่อสกัดข้อมูลรายชื่อจังหวัด และประเภทของยานพาหนะที่เกิดเหตุใช้เป็นเงื่อนไขหลัก สำหรับกำหนดเงื่อนไขรองในการค้นหาและระบุจำนวนของยานพาหนะจาก

ส่วนรายละเอียดของรายงานอุบัติเหตุ และ 2) เมื่อเงื่อนไขแรกเสร็จสิ้น จะค้นหาและระบุตำแหน่งของข้อความด้วยเรกูลาร์ เอ็กซ์เพรชชันร่วมกับการรู้จำชื่อเอนทิตี้อีกครั้ง เพื่อหาวันที่เกิดเหตุ จำนวนผู้บาดเจ็บ และจำนวนผู้เสียชีวิต จากนั้นจะกรองข้อมูลอีกครั้ง เพื่อให้ได้เฉพาะข้อมูลที่เป็นตัวเลข สำหรับคำนวณหาผลรวม และนำเสนอสารสนเทศด้วยภาพข้อมูลแก่ผู้ใช้

4.5 มอดูลการนำเสนอสารสนเทศ (Information Presentation Module) ด้วยภาพข้อมูลโดยใช้ Google Charts [15] ที่ทำงานผ่านส่วนต่อประสานโปรแกรมประยุกต์ (Application Program Interface: API) [16] โดยจะส่งข้อมูลไปประมวลผลที่เครื่องแม่ข่าย และรับข้อมูลกลับมาแสดงผลบนเว็บไซต์ในลักษณะเว็บแอปพลิเคชัน (Web Application) ที่ติดต่อปฏิสัมพันธ์กับผู้ใช้ผ่านเว็บเบราว์เซอร์

5. ผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อสกัดข้อมูลรายงานอุบัติเหตุทางถนน และทดสอบความสามารถในการนำเสนอสารสนเทศที่เกี่ยวข้อง ผลการวิจัยแบ่งเป็น 2 ส่วนดังนี้

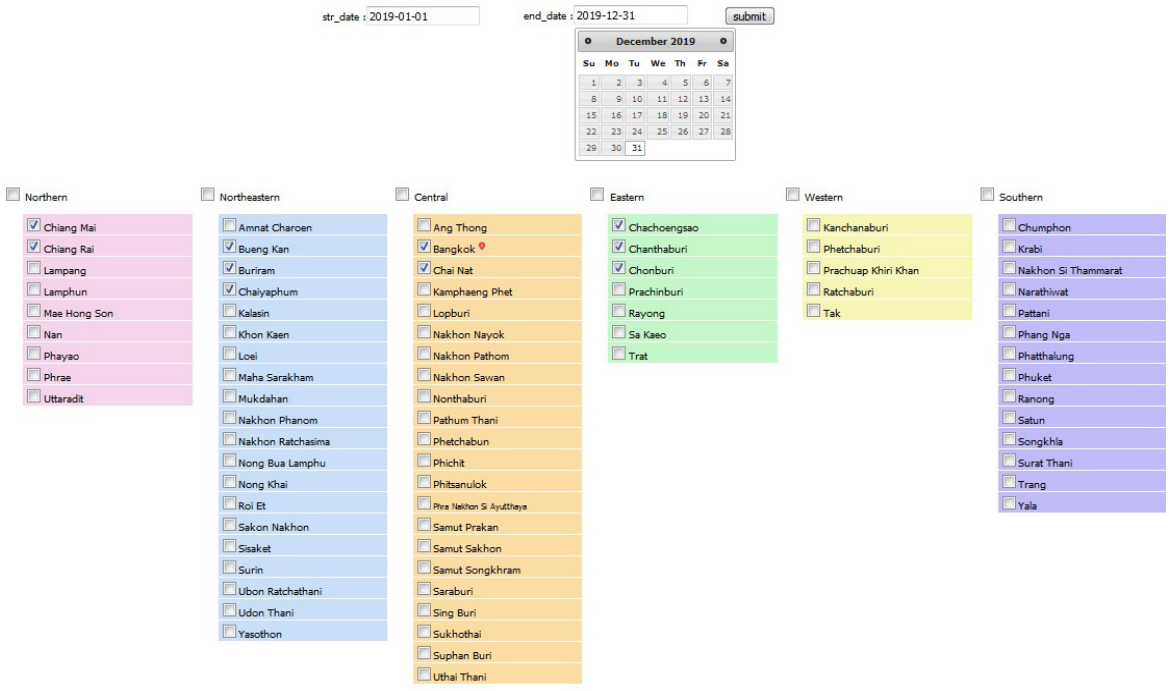
5.1 ผลการสกัดข้อมูลรายงานอุบัติเหตุทางถนน เริ่มจากผู้วิจัยนำเข้าข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่เข้าสู่ฐานข้อมูล จากนั้นจะคิวรี (Query) ส่วนหัวของรายงาน และส่วนรายละเอียดของรายงาน เพื่อสกัดข้อมูลตามกระบวนการที่ได้ออกแบบและพัฒนาไว้ในลักษณะวนลูป ดังรูปที่ 4

จากรูปที่ 4 ผลลัพธ์การสกัดข้อมูลจังหวัด “Buri Ram” ที่รูปแบบไม่เป็นไปตามมาตรฐาน แต่ผ่านกระบวนการเปรียบเทียบความคล้ายคลึงกันของอักขระด้วยเทคนิคการวัดระยะทางเลเวนชเตย์น จึงได้ผลลัพธ์เป็น “Buriram” เช่นเดียวกับวันที่จากการสกัดข้อมูล “06092019” ที่ปรับรูปแบบตามหลักสากลเป็น “2019-09-06” เพื่อนำไปคิวรี่ หาผลลัพธ์ช่วงระหว่างวันด้วยภาษาเอสคิวแอล (Structured Query Language: SQL) ขณะที่ผลลัพธ์จากการสกัดข้อมูลในประเด็นอื่น ๆ ผลลัพธ์เป็นไปตามกระบวนการที่ได้ออกแบบไว้

กระบวนการนำเสนอสารสนเทศ เริ่มจากรับข้อมูลเงื่อนไขในการค้นหาจากผู้ใช้ แบ่งเป็นการค้นหาจากระหว่างวันเริ่มต้นและวันสิ้นสุด ร่วมกับการค้นหาด้วยรายชื่อจังหวัด ดังรูปที่ 5 เมื่อการสกัดข้อมูลเสร็จสิ้นจะแสดงสถิติการเกิดอุบัติเหตุรายจังหวัดด้วยแผนที่ประเทศไทย ในการแสดงผลเปรียบเทียบสถิติการเกิดอุบัติเหตุ จำนวนผู้บาดเจ็บ และจำนวนผู้เสียชีวิตแต่ละจังหวัด และแสดงผลแยกตามประเภทของยานพาหนะที่เกิดอุบัติเหตุด้วยแผนภูมิวงกลม

Code	#ContentPlaceHolder1_lb_title	#section .clearfix	Extract Data
002711	Accident Pickup Car Collide with Sedan Car 6 Motorcycle 1 Death a lot of Injurie Nangrong District Buri Ram 06092019 0230 am	Accident Pickup Car Collide with Sedan Car 6 Motorcycle 1 Death a lot of Injurie Nangrong District Buri Ram 06092019 0230 am Accident Notification Centre Road Accident Victims Protection Co Ltd Department Disaster Prevention Mitigation Accident Report Notification EAccident 620010013913 Source Safety Radio Center Department of Disaster Prevention Mitigation Road Accident Victims Protection Company Limited Buri Ram branch Date Claim Friday September 6th 2019 0615 am Date Accident Friday September 6th 2019 0230 am Accident Point Nangrongchamni road Nangrong Nangrong District Buri Ram Accident Condition 4 lane straightway 22 Weather Condition usual Responsibility area Nangrong Police Station Tel 044631555 Enquiry officer Case owners Tel Accident scene Accident pickup car collide sedan car 6 motorcycle Photo Siam Buri Ram Rescue Siam Nangrong Motor Vehicle involve accident 1 Pickup car Isuzu DMax white color license plate registration xxxxxx Ayutthaya 2 Sedan car Honda City white color license plate registration xxxxxx Buri Ram 3 Motorcycle Honda Scoopy color license plate registration xxxxxx Buri Ram 4 Motorcycle Honda Click color license plate registration xxxxxxx Buri Ram 5 Motorcycle Honda Wave color license plate registration xxxxxx Buri Ram 6 Motorcycle color license plate registration 7 Motorcycle color license plate registration 8 Motorcycle color license plate registration List death 1 Death Male List injured 8 Injurie Male 4 Female 4 Injurie Identification Cause accident The real causes accident being investigated	Province: Buriram Date: 2019-09-06 Motor Vehicle(s): [6] MOTORCYCLE [1] CAR [1] PICK UP Death(s) 1 Injured(s) 8

รูปที่ 4 ผลลัพธ์จากการสกัดข้อมูลรายงานอุบัติเหตุทางถนน

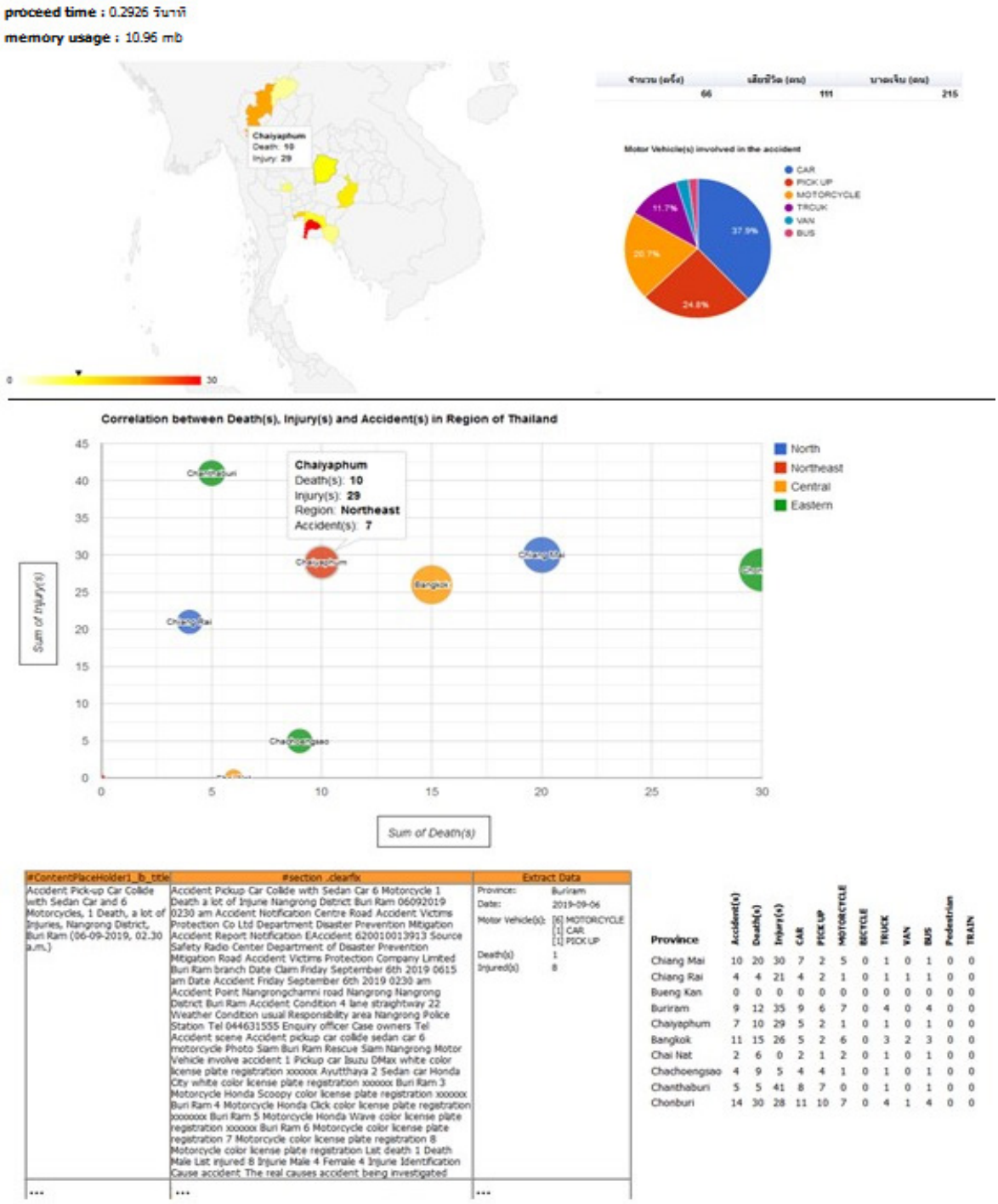


รูปที่ 5 การรับข้อมูลเงื่อนไขในการนำเสนอสารสนเทศรายงานอุบัติเหตุทางถนน

งานวิจัยนี้ใช้หลักการแสดงสีแบบไล่ลำดับ (Sequential Color) [17] โดยเลือกใช้สีขาว สีเหลือง สีส้ม และสีแดง ในการแสดงผล โดยไล่ลำดับความเข้มของสีจากน้อยที่สุดไปยังมากที่สุดตามปริมาณการเกิดอุบัติเหตุ ร่วมกับแสดงความสัมพันธ์ของการเกิดอุบัติเหตุของแต่ละจังหวัดเปรียบเทียบในระดับภูมิภาค ตลอดจนผลการสกัดข้อมูลแต่ละรายการดังรูปที่ 6 และปฏิสัมพันธ์กับผู้ใช้เมื่อนำเคอเซอร์ไปวางบนตามตำแหน่งที่ต้องการ

5.2 ผลการทดสอบความสามารถในการนำเสนอสารสนเทศรายงานอุบัติเหตุทางถนนแบ่งเป็น 3 ส่วน ดังนี้

5.2.1 ผลการทดสอบด้านประสิทธิภาพ (Efficiency) โดยวัดสมรรถนะของการสกัดข้อมูลจากระยะเวลาที่ใช้ (Proceed Time) และทรัพยากรระบบที่ใช้ (Memory Usage) ในการสกัดข้อมูล กำหนดสภาพแวดล้อมในการทดสอบด้วยเครื่องคอมพิวเตอร์ติดตั้งระบบปฏิบัติการไมโครซอฟต์ วินโดวส์เซเว่น (Microsoft Windows 7) หน่วยประมวลผลอินเทล คอร์ไอไฟว์ (Intel Core i5) ความเร็ว 2.30 กิกะเฮิรตซ์ (GHz) หน่วยความจำ 4 กิกะไบต์ (GB) และพื้นที่จัดเก็บข้อมูล 500 กิกะไบต์ เชื่อมต่ออินเทอร์เน็ตแบบไร้สาย ประมวลผลผ่านเว็บแอปพลิเคชันด้วยการเขียนโปรแกรมภาษา PHP ร่วมกับฐานข้อมูล MySQL จากรูปที่ 5 ผู้วิจัยทำการกำหนดเงื่อนไขในการทดสอบด้านประสิทธิภาพของระบบ โดยเลือกวันที่เริ่มต้นเป็น “2016-01-01” และวันสิ้นสุดเป็น “2020-03-31” ร่วมกับเลือกแสดงข้อมูลทั้ง 77 จังหวัดตามจำนวนจังหวัดของประเทศไทย ดำเนินการซ้ำ 3 ครั้งแล้วหาค่าเฉลี่ยเพื่อลดความคลาดเคลื่อนให้อยู่ในระดับที่ยอมรับได้ พบว่ามีผลลัพธ์รายการอุบัติเหตุที่แสดงผลทั้งสิ้น 1,353 รายการ ใช้เวลาเฉลี่ย 14.45 วินาทีต่อครั้ง และใช้ทรัพยากรระบบเฉลี่ย 49.20 เมกะไบต์ (MB) ต่อครั้ง



Province	Accident(s)	Death(s)	Injury(s)	CAR	PICK UP	MOTORCYCLE	BICYCLE	TRUCK	VAN	BUS	Pedestrian	TRAIN
Chiang Mai	10	20	30	7	2	5	0	1	0	1	0	0
Chiang Rai	4	4	21	4	2	1	0	1	1	1	0	0
Bueng Kan	0	0	0	0	0	0	0	0	0	0	0	0
Buriram	9	12	35	9	6	7	0	4	0	4	0	0
Chaiyaphum	7	10	29	5	2	1	0	1	0	1	0	0
Bangkok	11	15	26	5	2	6	0	3	2	3	0	0
Chai Nat	2	6	0	2	1	2	0	1	0	1	0	0
Chachoengsao	4	9	5	4	4	1	0	1	0	1	0	0
Chanthaburi	5	5	41	8	7	0	0	1	0	1	0	0
Chonburi	14	30	28	11	10	7	0	4	1	4	0	0

รูปที่ 6 ผลการนำเสนอสารสนเทศแบบภาพข้อมูล

5.2.2 ผลการทดสอบความสามารถในการนำเสนอสารสนเทศรายงานอุบัติเหตุทางถนนด้านประสิทธิผลของการค้นคืนข้อมูล (Retrieval Effectiveness) ของระบบที่พัฒนาขึ้น โดยทำการวัดความสามารถในการค้นคืนสารสนเทศแบบประเมินประสิทธิภาพเอกสารที่ถูกเลือก แต่มิได้เรียงลำดับความคล้ายคลึง [18] ประกอบด้วย 1) ค่าความถูกต้อง (Precision) ที่บอกถึงประสิทธิผลในการค้นคืนเอกสารโดยดูจากอัตราส่วนของจำนวนเอกสารที่ถูกต้องจากเอกสารที่ถูกเลือกมาทั้งหมด 2) ค่าความครบถ้วน (Recall) หมายถึง ประสิทธิผลของการค้นคืนเอกสารโดยดูจากอัตราส่วนจำนวนเอกสารที่ถูกต้องที่เลือกมาต่อจำนวนเอกสารที่ถูกต้องทั้งหมดที่อยู่ในคอลเลกชัน (Collection) และ 3) ค่าประสิทธิภาพโดยรวม (F-measure) หรือ F1 score คือการรวมค่าความถูกต้องตาม 1) และค่าความครบถ้วนเข้าด้วยกันโดยหาค่าเฉลี่ยที่ดีที่สุด (Weighted Harmonic Mean) การวิจัยนี้ทดสอบความสามารถในการนำเสนอสารสนเทศรายงานอุบัติเหตุทางถนนด้านประสิทธิผลของการค้นคืนข้อมูลด้วยเงื่อนไขจังหวัด และวันที่ โดยเลือกเงื่อนไขทีละครั้ง จำนวน 77 ครั้งตามจำนวนจังหวัดของประเทศไทย เมื่อครบรอบจะทำการทดสอบอีกครั้งจนครบ 3 รอบ แบบผลลัพธ์เป็นอิสระต่อกัน เพื่อนำผลการทดสอบมาหาค่าเฉลี่ยจนค่าความคลาดเคลื่อนอยู่ในระดับที่ยอมรับได้ พบว่า ระบบที่พัฒนาขึ้นสามารถค้นคืนรายงานอุบัติเหตุได้ตรงตามเงื่อนไขการค้นหา โดยมีค่าความถูกต้องเท่ากับ 0.87 ค่าความครบถ้วน เท่ากับ 0.85 และค่าประสิทธิภาพโดยรวมเท่ากับ 0.86 ซึ่งทั้ง 3 ค่ามีผลการทดสอบมากกว่าร้อยละ 80 ดังนั้นผลการทดสอบเป็นไปตามสมมติฐานการวิจัย

5.2.3 ผลการทดสอบความสามารถในการนำเสนอสารสนเทศการสกัดข้อมูลรายงานอุบัติเหตุทางถนน ตามหลักการใช้งานได้ของเว็บไซต์ (Web Usability) ที่มีพื้นฐานจากการบูรณาการหลายศาสตร์เข้าด้วยกัน เช่น วิทยาการคอมพิวเตอร์ จิตวิทยา พฤติกรรมศาสตร์ และการออกแบบ [19] เพื่อเพิ่มประสิทธิภาพในการปฏิสัมพันธ์ระหว่างมนุษย์และคอมพิวเตอร์ นอกจากนี้ยังได้รับแนวคิดมาจากการออกแบบโดยเน้นการตอบสนองผู้ใช้เป็นหลัก (User-Centered Designed) และการยศาสตร์ (Ergonomics) โดยทั้งสองแนวคิดให้ความสำคัญกับเทคนิค ขั้นตอน วิธีการและกระบวนการที่ช่วยให้สามารถออกแบบผลิตภัณฑ์ที่เหมาะสมกับมนุษย์ทั้งทางร่างกายและจิตใจ เพื่อเพิ่มประสิทธิภาพและความปลอดภัยที่วัดคุณภาพจากความง่ายในการใช้งานระบบ โดยเกี่ยวข้องกับระยะเวลาเรียนรู้การใช้งาน ประสิทธิภาพในการทำงานของระบบ ความยากง่ายในการจดจำ และข้อผิดพลาดของระบบ ตลอดจนความพึงพอใจของผู้ใช้ มีหลายทฤษฎีในการกำหนดเป้าหมายการใช้งานได้ โดย ISO [20] ได้กำหนดองค์ประกอบการใช้งานของระบบเทคโนโลยีสารสนเทศในมาตรฐานชุด ISO 9241-11 เป็นมาตรฐานทางการยศาสตร์ที่เกี่ยวกับการทำงานของมนุษย์ร่วมกับคอมพิวเตอร์ 3 ด้านดังนี้ คือ 1) ด้านประสิทธิผล (Effectiveness) คือความถูกต้องของเป้าหมายที่ผู้ใช้บรรลุในขอบเขตของระบบ 2) ด้านประสิทธิภาพ (Efficiency) คือ ปริมาณทรัพยากรที่ระบบใช้เพื่อให้บรรลุเป้าหมายอย่างสมบูรณ์ และ 3) ด้านความพึงพอใจ (Satisfaction) คือความสะดวกสบายและความยอมรับของผู้ใช้ในการใช้งานระบบ และเมื่อพิจารณาเป้าหมายการใช้งานได้ที่กำหนดโดยนักวิจัยหลายท่านตั้งแต่อดีตจนถึงปัจจุบัน ได้แก่ Shackel และ Richardson [21], Preece [22], Dix และคณะ [19], Shneiderman และ Plaisant [23], Brinck และคณะ [24] ตลอดจน Rubin และ Chisnell [25] จัดแบ่งประเภทเป้าหมายของการใช้งานได้เป็น 4 ด้าน คือ 1) ความมีประโยชน์ (Usefulness) ของระบบที่ช่วยผู้ใช้ให้บรรลุผลสำเร็จของงานและทำงานได้ถูกต้อง 2) ประสิทธิผลและประสิทธิภาพ (Effectiveness and Efficiency) คือความง่ายและความรวดเร็วในการทำงานให้บรรลุผลสำเร็จ ช่วยป้องกันข้อผิดพลาดในการทำงาน และการแก้ไขเมื่อเกิดข้อผิดพลาดขึ้น 3) การเรียนรู้ (Learnability) คือความง่ายที่ผู้ใช้สามารถเรียนรู้วิธีการใช้งานระบบ หรือหากผู้ใช้ไม่ได้ใช้งานระบบระยะเวลาหนึ่ง เมื่อกลับมาใช้งานอีกครั้งยังคงจดจำวิธีใช้งานได้อย่างรวดเร็ว และ 4) ความพึงพอใจของผู้ใช้ (Likability) ที่ระบบสามารถทำให้ผู้ใช้เกิดความพึงพอใจในการใช้งานได้เพียงใด สอดคล้องกับ Nielsen [26] ที่กำหนดเป้าหมายของการใช้งานในแนวทางเดียวกันแต่แยกการจดจำ (Memorability) ออกมาจากองค์ประกอบด้านการเรียนรู้ เพื่อประเมินว่าจะใช้เวลาเรียนรู้การใช้งานนานเพียงใดเมื่อผู้ใช้กลับไปใช้งานอีกครั้ง หลังจากไม่ได้ใช้งานเป็นระยะเวลาหนึ่ง

ผลการสังเคราะห์เอกสารที่เกี่ยวข้อง พบว่า องค์ประกอบการใช้งานที่จำเป็นของทุกระบบมีองค์ประกอบหลักใกล้เคียงกัน จึงนำมาเปรียบเทียบได้ดังตารางที่ 2 เมื่อนำมากำหนดองค์ประกอบความสามารถในการนำเสนอสารสนเทศการสกัดข้อมูลรายงานอุบัติเหตุทางถนน ประกอบด้วย 5 องค์ประกอบ ได้แก่ ด้านประสิทธิภาพ (Efficiency) ด้านประสิทธิผล (Effectiveness) ด้านการเรียนรู้ (Learnability) ด้านความยืดหยุ่น (Flexibility) และด้านความพึงพอใจ (Satisfaction) รวมทั้งสิ้น 20 ข้อ

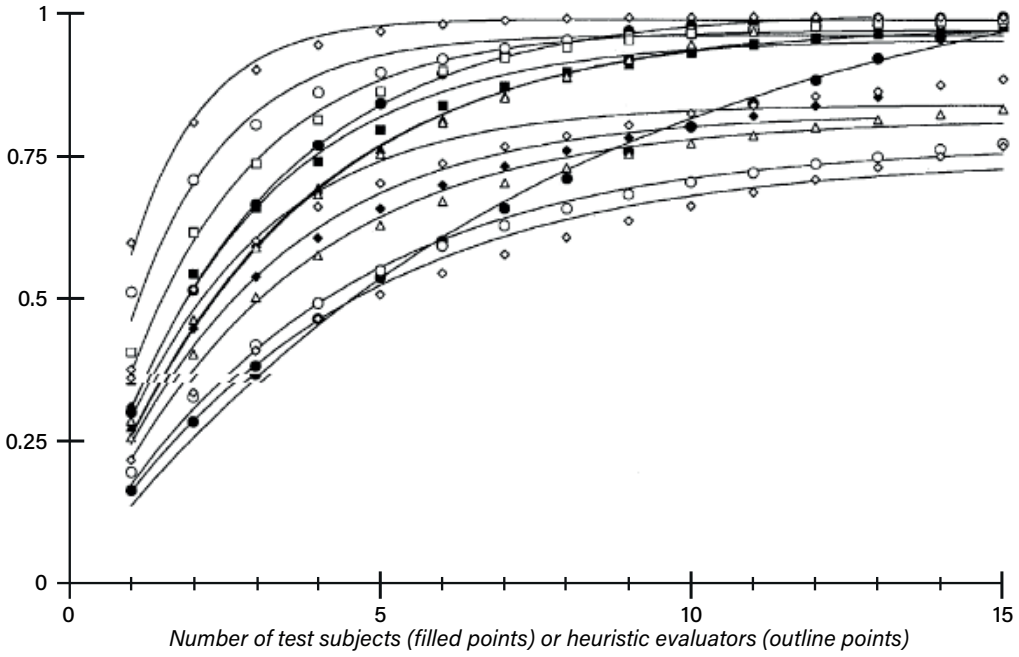
ผู้วิจัยได้นำแบบประเมินที่สร้างขึ้นสำหรับการวิจัยไปทำการทดสอบคุณภาพเครื่องมือ 2 ส่วน คือ 1) ค่าความเที่ยงตรงของเนื้อหา (Validity) โดยการนำแบบประเมินที่ผู้วิจัยสร้างขึ้นเสนอต่อผู้เชี่ยวชาญพิจารณาโดยกำหนดเกณฑ์ในการคัดเลือกผู้เชี่ยวชาญ (Experts) หรือผู้มีประสบการณ์ (Specialist) หรือผู้ทรงคุณวุฒิ [27] หมายถึงผู้มีทักษะหรือผู้ที่เคยลงมือปฏิบัติงานใด ๆ จนได้ผลดีเป็นที่ประจักษ์ชัด หรือผู้ที่มีทัศนเชิงวิทยาศาสตร์ในวิชาชีพที่มีพยานประจักษ์ชัดว่าได้ปฏิบัติการในวิชาชีพจนได้รับผลสำเร็จดี หรือเป็นผู้ที่ได้รับการฝึกฝน มีความรู้ความชำนาญเป็นเลิศในสาขาวิชานั้น ๆ หรือเป็นผู้ที่มีความรู้เป็นพิเศษในสาขาที่สนใจ งานวิจัยนี้กำหนดเกณฑ์การคัดเลือกผู้เชี่ยวชาญที่เหมาะสมสอดคล้องกับเครื่องมือหรือเรื่องที่จะทำวิจัย โดยพิจารณาในประเด็นเหล่านี้ประกอบกัน ได้แก่ 1) ผู้ที่มีความรู้ ความเข้าใจในสาขาวิชานั้น 2) ผู้ที่มีผลงานเกี่ยวข้องในสาขาวิชานั้น 3) ผู้ที่มีวุฒิทางการศึกษาจนเป็นที่เชื่อถือหรือยอมรับในสาขานั้น 4) ผู้ที่มีทักษะเฉพาะตัวในสาขานั้น โดยได้รับเกียรติบัตรหรือปริญญาบัตรกิตติมศักดิ์จากสถาบันที่เชื่อถือได้ 5) ผู้ที่มีประสบการณ์ในการปฏิบัติงานเกี่ยวกับสาขานั้นสูง และ 6) ผู้ที่มีความตั้งใจจริงและเสียสละเวลาในการตรวจสอบเครื่องมืออย่างเต็มที่ งานวิจัยนี้กำหนดผู้เชี่ยวชาญ 3 ท่านพิจารณาจากค่าความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์หรือเนื้อหาในการวิจัย ประกอบด้วยผู้ทรงคุณวุฒิด้านเทคโนโลยีสารสนเทศ ผู้ทรงคุณวุฒิด้านสถิติ และผู้ปฏิบัติงานด้านสารสนเทศจราจรช่วยตรวจสอบความถูกต้องและความครอบคลุมของเนื้อหาที่ต้องการศึกษา ทั้งนี้ผู้วิจัยได้วิเคราะห์ดัชนีความสอดคล้อง (Index of Item Objective Congruence: IOC) ได้เท่ากับ 0.86 และเลือกข้อคำถามที่มีค่า IOC มากกว่า 0.5 มาใช้เป็นข้อคำถาม [28] และ 2) ความเชื่อมั่น (Reliability) ผู้วิจัยได้นำแบบประเมินที่สร้างขึ้นมาจำนวน 30 ชุดไปทำการทดสอบกับกลุ่มตัวอย่างเพื่อตรวจสอบว่าคำถามสามารถสื่อความหมายตรงตามความต้องการและมีความเหมาะสมหรือไม่ จากนั้นจึงนำมาทดสอบความเชื่อมั่นของแบบประเมินโดยหาค่าสัมประสิทธิ์แอลฟา (Cronbach's Alpha Coefficient) [29] มีค่าระดับความเชื่อมั่นของแบบประเมินเท่ากับ 0.82 ซึ่งถือได้ว่าอยู่ในระดับดี [30] หมายถึงแบบประเมินมีความน่าเชื่อถือและสามารถนำไปทดสอบกับกลุ่มตัวอย่างจริงได้

ตารางที่ 2 องค์ประกอบการใช้งาน

Usability Components	[20]	[21]	[22]	[19]	[23]	[24]	[25]	[26]	In this Article
Usefulness							✓		
Efficiency	✓		✓		✓	✓	✓	✓	✔
Effectiveness	✓	✓	✓	✓	✓	✓	✓	✓	✔
Learnability		✓	✓	✓	✓	✓	✓	✓	✔
Memorability			✓		✓	✓		✓	
Flexibility		✓		✓					✔
Satisfaction	✓	✓			✓	✓		✓	✔

การกำหนดขนาดของกลุ่มตัวอย่างนั้น Nielsen และ Landauer [31] ได้เสนอจำนวนผู้ทดสอบความสามารถในการใช้งานไว้ว่า 5 คน จะพบปัญหาในระบบได้ 85% และถ้าผู้ทดสอบ 15 คนหรือมากกว่าจะพบปัญหาทั้งหมดภายในระบบ ดังรูปที่ 7 งานวิจัยนี้กำหนดกลุ่มตัวอย่างจำนวน 30 คน เพื่อพบปัญหาในระบบ 100%

อย่างไรก็ดี แม้ผลการวิจัยที่เลือกกลุ่มตัวอย่างแบบเฉพาะเจาะจงนั้นจะไม่สามารถสรุปอ้างอิงไปสู่ประชากรโดยทั่วไปได้ แต่งานวิจัยนี้เพิ่มความน่าเชื่อถือของผลลัพธ์ผ่านการคัดเลือกผู้เชี่ยวชาญด้วยการใช้เทคนิคเดลฟาย (Delphi Technique) [33] ซึ่งขนาดของกลุ่มตัวอย่างไม่ได้ขึ้นอยู่กับการวิเคราะห์ทางสถิติ แต่ขึ้นอยู่กับความเป็นพลวัตของกลุ่ม



รูปที่ 7 จำนวนปัญหาที่พบในระบบต่อจำนวนผู้ทดสอบ [31]

งานวิจัยนี้ใช้การสุ่มตัวอย่างโดยไม่ใช้ความน่าจะเป็น (Non-Probability Sampling) ในการเก็บข้อมูล เนื่องจากสามารถเก็บข้อมูลจากกลุ่มตัวอย่างจำนวนมากได้ในระยะเวลาจำกัด จากนั้นใช้วิธีการเลือกกลุ่มตัวอย่างแบบเฉพาะเจาะจง (Purposive Sampling) [32] เป็นการเลือกกลุ่มตัวอย่างให้สอดคล้องเหมาะสมกับวัตถุประสงค์การวิจัยที่กำหนดไว้ เนื่องจากงานวิจัยนี้มีวัตถุประสงค์เพื่อสกัดข้อมูลรายงานอุบัติเหตุทางถนน และทดสอบความสามารถในการนำเสนอสารสนเทศที่เกี่ยวข้อง จึงกำหนดเงื่อนไขและคุณลักษณะของผู้ใช้ที่สุ่มมาเป็นตัวอย่าง ตามบริบทในการใช้งานสารสนเทศการสกัดข้อมูลรายงานอุบัติเหตุทางถนน ฉันทามติระหว่างผู้เชี่ยวชาญ ควรมีผู้เชี่ยวชาญ 10-18 คน [34] งานวิจัยนี้สิ่งที่ผู้วิจัยจะต้องคำนึงถึงในการเลือกกลุ่มผู้เชี่ยวชาญ ได้แก่ ความสามารถของกลุ่มผู้เชี่ยวชาญ ความร่วมมือของผู้เชี่ยวชาญ จำนวนผู้เชี่ยวชาญ และวิธีการเลือกสรรผู้เชี่ยวชาญตามบริบทในการใช้งานสารสนเทศการสกัดข้อมูลรายงานอุบัติเหตุทางถนนออกเป็น 3 กลุ่ม กลุ่มละ 10 คน ได้แก่ กลุ่มผู้เชี่ยวชาญด้านคอมพิวเตอร์ กลุ่มผู้ทำหน้าที่ข้อมูลจราจร และกลุ่มผู้ใช้งานทั่วไป รวมทั้งสิ้น 30 คน

กำหนดการทดสอบรูปแบบกล่องดำ (Blackbox Testing) มีขั้นตอนดังนี้ ก่อนการทดสอบผู้วิจัยอธิบายรายละเอียดวัตถุประสงค์ และวิธีการทดสอบให้กลุ่มเป้าหมาย

เข้าใจชัดเจน ระหว่างการทดสอบให้กลุ่มเป้าหมายใช้งานเว็บแอปพลิเคชันที่พัฒนาขึ้นอย่างอิสระ โดยผู้วิจัยสังเกตตลอดระยะเวลาการทดสอบ ภายหลังทดสอบการใช้งานแล้วให้กลุ่มเป้าหมายประเมินผลความสามารถในการนำเสนอสารสนเทศผ่านแบบประเมินระดับการวัดคุณภาพตามมาตรวัดลิเคิร์ท [35] 5 ระดับ (5 - Point Likert Scale) ได้แก่

1 หมายถึง ระดับน้อยที่สุด
2 หมายถึง ระดับน้อย
3 หมายถึง ระดับปานกลาง
4 หมายถึง ระดับมาก
5 หมายถึง ระดับมากที่สุด

จากนั้นจึงใช้กระบวนการทางสถิติประมวลและสรุปผลรายละเอียดดังตารางที่ 3 แปลผลค่าเฉลี่ยแต่ละองค์ประกอบ [36] ตามเกณฑ์ 5 ระดับ ได้แก่

4.51 – 5.00 หมายถึง ระดับมากที่สุด
3.51 – 4.50 หมายถึง ระดับมาก
2.51 – 3.50 หมายถึง ระดับปานกลาง
1.51 – 2.50 หมายถึง ระดับน้อย
1.00 – 1.50 หมายถึง ระดับน้อยที่สุด

ตารางที่ 3 ผลการทดสอบความสามารถในการนำเสนอสารสนเทศการสกัดข้อมูลรายงานอุบัติเหตุทางถนน

ข้อ	ประเด็นคำถามการใช้งาน	$\bar{X}$	S.D.	แปลผล	Beta	Tolerance	VIF
ด้านประสิทธิภาพ (Efficiency)							
1.	ความรวดเร็วในการนำเสนอผลการสกัดข้อมูล (X_1)	3.23	0.74	ปานกลาง	.302	.258	3.879
2.	ความรวดเร็วในการนำเสนอภาพข้อมูล (X_2)	3.57	0.67	มาก	.286	.341	2.933
3.	ระยะเวลาการตอบสนองของภาพข้อมูลเมื่อถูกกระทำ (X_3)	3.73	0.60	มาก	.320	.301	3.322
	ค่าเฉลี่ย	3.51	0.64	มาก			
ด้านประสิทธิผล (Effectiveness)							
4.	ความถูกต้องของการแสดงผลข้อมูลตามเงื่อนไขการค้นหา (X_4)	3.37	0.78	ปานกลาง	.314	.198	5.288
5.	ความถูกต้องของผลการสกัดข้อมูลรายงานอุบัติเหตุ (X_5)	3.30	0.79	ปานกลาง	.292	.226	4.428
6.	ความถูกต้องของข้อมูลเมื่อนำเสนอในรูปแบบภาพข้อมูล (X_6)	3.37	0.70	ปานกลาง	.316	.463	2.159
7.	ความถูกต้องของการจัดวางรายชื่อจังหวัดเพื่อการค้นหา (X_7)	3.37	0.70	ปานกลาง	.292	.568	1.760
8.	ความถูกต้องการของการใช้สีในการแสดงความแตกต่างของข้อมูลแต่ละประเภท (X_8)	3.77	0.70	มาก	.294	.401	2.492
	ค่าเฉลี่ย	3.43	0.73	ปานกลาง			

ตารางที่ 3 ผลการทดสอบความสามารถในการนำเสนอสารสนเทศการสกัดข้อมูลรายงานอุบัติเหตุทางถนน (ต่อ)

ข้อ	ประเด็นคำถามการใช้งาน	$\bar{X}$	S.D.	แปลผล	Beta	Tolerance	VIF
ด้านการเรียนรู้ (Learnability)							
9.	ความสามารถในการเรียนรู้การใช้ระบบได้เอง โดยไม่ต้องใช้ผู้เชี่ยวชาญช่วยเหลือ (X_9)	3.93	0.53	มาก	.220	.274	3.643
10.	ความสามารถในการใช้งานระบบได้อย่างรวดเร็ว (X_{10})	4.00	0.49	มาก	.210	.367	2.728
11.	ความสามารถในการทำความเข้าใจกับข้อมูลที่นำเสนอในลักษณะภาพข้อมูลได้ง่าย (X_{11})	4.33	0.41	มาก	.189	.231	4.333
12.	ความสามารถในการทำความเข้าใจกับกระบวนการได้มาซึ่งผลลัพธ์ได้ด้วยตนเอง (X_{12})	3.87	0.60	มาก	.240	.218	4.577
	ค่าเฉลี่ย	4.03	0.52	มาก			
ด้านความยืดหยุ่น (Flexibility)							
13.	การนำเสนอภาพข้อมูลมีการตอบสนองและปรับเปลี่ยนได้ตามที่ผู้ใช้ต้องการ (X_{13})	3.77	0.64	มาก	.220	.299	3.349
14.	ความเหมาะสมของหน้าเว็บไซต์เมื่อแสดงผลบนอุปกรณ์ที่มีขนาดต่างกัน (X_{14})	3.80	0.73	มาก	.210	.395	2.533
15.	ความเหมาะสมของหน้าเว็บไซต์เมื่อแสดงผลบนเว็บเบราว์เซอร์ที่แตกต่างกัน (X_{15})	3.23	0.67	ปานกลาง	.189	.410	2.442
	ค่าเฉลี่ย	3.60	0.70	มาก			
ด้านความพึงพอใจ (Satisfaction)							
16.	ระบบมีรูปแบบการจัดวางองค์ประกอบที่ชัดเจน (X_{16})	4.13	0.48	มาก	.207	.332	3.015
17.	สีสันที่นำมาใช้แสดงผลมีความโดดเด่น สวยงาม (X_{17})	3.90	0.46	มาก	.202	.244	4.092
18.	ขนาดของรูปภาพ และข้อความ มีความเหมาะสม (X_{18})	3.93	0.53	มาก	.220	.303	3.301
19.	ความเหมาะสมของการเลือกใช้ภาพข้อมูลแต่ละประเภทในการนำเสนอ (X_{19})	4.13	0.56	มาก	.229	.316	3.167
20.	ความเหมาะสมของการเลือกใช้สีในการแสดงระดับความรุนแรงของอุบัติเหตุ (X_{20})	3.90	0.58	มาก	.236	.213	4.705
	ค่าเฉลี่ย	4.00	0.52	มาก			
	ค่าเฉลี่ยรวมทุกองค์ประกอบ	**3.73**	**0.63**	**มาก**			

จากตารางที่ 3 ผลการทดสอบความสามารถในการนำเสนอสารสนเทศการสกัดข้อมูลรายงานอุบัติเหตุทางถนนภาพรวมอยู่ในระดับดี ($\bar{X}$= 3.73) เมื่อพิจารณาแต่ละด้าน พบว่าทุกด้านมีค่าเฉลี่ยอยู่ในระดับมากขึ้นไป โดยองค์ประกอบที่เป็นจุดเด่น คือ ด้านการเรียนรู้ และด้านความพึงพอใจ ซึ่งมีค่าเฉลี่ยที่ใกล้เคียงกัน ($\bar{X}$ = 4.03 และ 4.00 ตามลำดับ) เมื่อพิจารณาในรายละเอียด พบว่า ประเด็นความสามารถในการทำความเข้าใจกับข้อมูลที่นำเสนอในลักษณะภาพข้อมูลได้ง่าย (X_{11}) จากองค์ประกอบด้านการเรียนรู้ มีผลการทดสอบสูงที่สุด ($\bar{X}$ = 4.33) ซึ่งผลการทดสอบมากกว่า 3.51 ดังนั้นผลการทดสอบเป็นไปตามสมมติฐานการวิจัย

อย่างไรก็ดี ระบบยังมีจุดที่ควรพัฒนาด้านประสิทธิผลอันจะเห็นได้จากผลการทดสอบในประเด็นความถูกต้องของการแสดงผลข้อมูลตามเงื่อนไขการค้นหา (X_4) และความถูกต้องของผลการสกัดข้อมูลรายงานอุบัติเหตุ (X_5) จากองค์ประกอบด้านประสิทธิผลอันมีที่มาจากกระบวนการสกัดข้อมูลรูปแบบเดียวกันนั้นมีผลการทดสอบแตกต่างกันเล็กน้อย ($\bar{X}$ = 3.37 และ 3.30 ตามลำดับ) แต่การกระจายตัวของข้อมูล (Standard Deviation: S.D.) กลับสูงกว่าประเด็นอื่น ๆ อย่างชัดเจน (S.D. = 0.78 และ 0.79 ตามลำดับ)

การวิเคราะห์ข้อมูลโดยใช้สถิติอนุมานโดยอาศัยเทคนิคการวิเคราะห์ความถดถอยพหุคูณ (Multiple Regression Analysis) ดังนั้นก่อนการวิเคราะห์ข้อมูล ผู้วิจัยได้ทำการตรวจสอบเงื่อนไขของการวิเคราะห์การถดถอยเชิงเส้น ดังนี้

1) การตรวจสอบความผิดปกติของข้อมูล (Outlier) ด้วยวิธีการวัดระยะห่างมาฮาลานอบิส (Mahalanobis Distance) โดยพิจารณาจากค่า p-value ต้องไม่น้อยกว่า 0.001 ($P < 0.001$) จึงถือว่าข้อมูลไม่ผิดปกติ [37] ซึ่งพบว่าค่า p ที่น้อยกว่า 0.001 แสดงว่าข้อมูลชุดนี้ไม่มีความผิดปกติของข้อมูล สามารถทำการวิเคราะห์ต่อได้

2) ตัวแปรตามและค่าความคลาดเคลื่อนเป็นตัวแปรที่มีการแจกแจงแบบปกติ (Test of Normality) ด้วยวิธีการทดสอบโคลโมโกรอฟ-สเมอร์นอฟ (Kolmogorov-Smirnov Test) [38] โดยมีระดับนัยสำคัญที่มากกว่า 0.05 ผลการทดสอบมีระดับนัยสำคัญหรือค่า Sig = 0.003 น้อยกว่า 0.05 แสดงว่าตัวแปรที่เป็นกลุ่มตัวอย่างมีการแจกแจงแบบไม่ปกติ แต่จากการทบทวนวรรณกรรมพบทฤษฎีแนวโน้มเข้าสู่ส่วนกลาง (Central Limit Theorem) ซึ่งระบุว่าสำหรับประชากรใด ๆ ถ้าเก็บตัวอย่างในจำนวนที่มากพอ การกระจายของค่าตัวอย่างดังกล่าวจะมีแนวโน้มใกล้เคียงกับการกระจายแบบธรรมชาติ (Normal Distribution) [39] ซึ่งทฤษฎีแนวโน้มเข้าสู่ส่วนกลางระบุว่าจำนวนของกลุ่มตัวอย่างที่ทำให้การแจกแจงเป็นแบบปกติ ควรมีมากกว่า 50 ตัวอย่าง [40] อย่างไรก็ดี กลุ่มตัวอย่างในงานวิจัยนี้เท่ากับ 30 ตัวอย่าง ผู้วิจัยจึงพิจารณาเลือกใช้ผลการทดสอบ Shapiro-Wilk Test พบว่า ผลการทดสอบ Sig. เท่ากับ 0.093 มากกว่า 0.05 ถือว่าตัวแปรตามและค่าความคลาดเคลื่อนเป็นตัวแปรที่มีการแจกแจงแบบปกติ

3) ค่าแปรปรวนของความคลาดเคลื่อนเป็นค่าคงที่ที่ไม่ทราบค่า (Homoscedasticity) โดยพิจารณาจากแผนภาพการกระจาย (Scatter Plot) หากค่าความคลาดเคลื่อนเปลี่ยนแปลงใกล้ 0 หรือมีการเปลี่ยนแปลงในช่วงแคบแสดงว่าค่าแปรปรวนของความคลาดเคลื่อนที่เกิดจากการพยากรณ์นั้นเป็นค่าคงที่ พบว่า ค่าความคลาดเคลื่อนส่วนใหญ่กระจายอยู่เหนือและใต้ระดับ 0 ซึ่งจากการกระจายตัวอยู่ในช่วงแคบ และอยู่ในแนวเส้นตรงไม่ว่าตัวแปรตามจะเปลี่ยนแปลงไปในทิศทางใด [41] ดังนั้นจึงสรุปว่าค่าความแปรปรวนของความคลาดเคลื่อนเป็นค่าคงที่

4) ค่าความคลาดเคลื่อนแต่ละค่ามีความเป็นอิสระต่อกัน (Autocorrelation) ตัวแปรอิสระต้องเป็นข้อมูลที่ไม่มีความสัมพันธ์ภายในตัวเอง โดยใช้ค่า Durbin-Watson ในการทดสอบว่าตัวแปรอิสระมีความสัมพันธ์ภายในตัวเองหรือไม่ หากมีค่าน้อยกว่า 1.5 และมากกว่า 2.5 [42] แสดงว่าเกิด Autocorrelation หรือตัวแปรอิสระมีความสัมพันธ์ภายในตัวเอง ซึ่งจะทำให้การคำนวณในสมการการวิเคราะห์ความถดถอยพหุคูณเชิงเส้นมีปัญหา จากการพิจารณาค่า Durbin-Watson พบว่า เท่ากับ 1.617 ซึ่งอยู่ระหว่าง 1.5 - 2.5 แสดงว่าตัวแปรอิสระที่นำมาใช้ในการทดสอบไม่มีความสัมพันธ์ภายในตัวเอง

5) ตัวแปรอิสระแต่ละตัวต้องไม่มีความสัมพันธ์กัน (Multicollinearity) โดยค่าสถิติสำหรับการตรวจสอบปัญหาภาวะร่วมเส้นตรงพหุคูณ ได้แก่ 1) ค่าความทนหรือความคลาดเคลื่อนยินยอม (Tolerance) หมายถึงความแปรปรวนในตัวแปรอิสระแต่ละตัวที่ไม่สามารถอธิบายได้ด้วยตัวแปรอิสระอื่น ๆ พบว่า ตัวแปรอิสระทั้งหมด มีค่า

Tolerance อยู่ระหว่าง 0.198 - 0.568 ดังตารางที่ 3 ซึ่งมากกว่า 0.19 ถือว่าผ่านเกณฑ์ [43] และ 2) และอัตราความแปรปรวนเฟ้อ (Variance Inflation Ratio: VIF) เนื่องจากค่าความทนมีความหมายตรงกันข้ามกับค่าสถิติ คือ ค่าความทนสูง หมายความว่ามีปัญหาภาวะร่วมเส้นตรงพหุต่ำ เป็นที่มาของการสร้างค่าสถิติ VIF เพื่อช่วยให้การแปลความหมายง่ายขึ้น กำหนดเกณฑ์ในการพิจารณา คือ ค่า VIF < 5.3 พบว่า มีค่า VIF อยู่ระหว่าง 1.760 - 5.288 ดังตารางที่ 3 ซึ่งน้อยกว่า 5.30 ถือว่าผ่านเกณฑ์ [43] สำหรับผลการวิเคราะห์ข้อมูลสรุปผลได้ว่าตัวแปรอิสระทุกตัวแปรในโมเดลการวิเคราะห์ถดถอยพหุคูณไม่มีปัญหาภาวะร่วมเส้นตรงพหุ

ต่อมาจึงนำตัวแปรอิสระทั้งหมดมาวิเคราะห์หาความสัมพันธ์ด้วยการวิเคราะห์ถดถอยเชิงเส้นแบบนำตัวแปรเข้าทั้งหมด (Enter Regression) เนื่องจากทั้งหมดเป็นตัวแปรเชิงปริมาณ และตั้งสมมติฐานว่าตัวแปรอิสระทุกตัวมีอิทธิพลต่อตัวแปรตามรายละเอียดดังตารางที่ 3 ผลการวิเคราะห์ถดถอยพหุคูณแบบนำตัวแปรทั้งหมดเข้าสู่สมการ โดยเปรียบเทียบค่าน้ำหนักของตัวแปรอิสระที่ส่งผลต่อตัวแปรตามจากค่า The Standardized beta (β) ที่มีช่วงระหว่างค่า 0 ถึง 1 หรือ 0 ถึง -1 ขึ้นกับทิศทางของความสัมพันธ์ พบว่า ทั้งหมดอยู่ระหว่าง .180 - .320 เท่านั้น นับว่าตัวแปรอิสระแต่ละตัวส่งผลต่อตัวแปรตามในระดับต่ำ

จากนั้นนำตัวแปรทั้งหมดเข้าวิเคราะห์อีกครั้งด้วยการคัดเลือกแบบลำดับขั้น (Stepwise Selection) เพื่อวิเคราะห์ว่าตัวแปรอิสระใดบ้างที่มีความสัมพันธ์กับตัวแปรตาม และแต่ละตัวมีอิทธิพลต่อผลการทดสอบแตกต่างกันอย่างไร กำหนดตัวแปรอิสระคือผลการทดสอบความสามารถในการนำเสนอสารสนเทศจากผู้ใช้จำนวน 20 ข้อ แทนด้วย X_1 - X_{20} และกำหนดผลการทดสอบความสามารถในการนำเสนอสารสนเทศโดยรวม ($\bar{x}$ = 3.73) เป็นตัวแปรตาม ดังตารางที่ 4 ผลการวิเคราะห์สมการถดถอยพหุคูณแบบลำดับขั้น เริ่มจากการเลือกตัวแปรอิสระที่มีสหสัมพันธ์กับตัวแปรตามสูงที่สุดเข้าสมการก่อน จากนั้นจะทดสอบตัวแปรที่ไม่ได้อยู่ในสมการว่าจะมีตัวทำนายตัวใดบ้างมีสิทธิ์เข้ามาอยู่ในสมการด้วยวิธีการคัดเลือกแบบก้าวหน้า (Forward Selection) และขณะเดียวกันก็จะทดสอบตัวแปรที่อยู่ในสมการด้วยว่าตัวแปรใดมีโอกาสที่จะถูกขจัดออกจากสมการด้วยวิธีการคัดเลือกแบบถอยหลัง (Backward Selection) โดยจะคัดเลือกผสมทั้งสองวิธีในทุกขั้นตอน กระทั่งไม่มีตัวแปรใดที่ถูกคัดออก และไม่มีตัวแปรใดที่จะถูกนำเข้าสมการอีก กระบวนการจะยุติและได้สมการถดถอยที่มีสัมประสิทธิ์การทำนายสูงสุด

ผลการวิเคราะห์สมการถดถอยพหุเพื่อหาความสัมพันธ์ระหว่างตัวแปรอิสระที่ส่งผลต่อผลการทดสอบความสามารถ

ตารางที่ 4 ผลการวิเคราะห์สมการถดถอยพหุคูณแบบลำดับขั้น

ตัวแปร / รูปแบบ	การดำเนินการ	B									R
		X_{19}	X_{11}	X_5	X_{15}	X_{10}	X_4	X_{12}	X_7	X_2	
1)	Entered	0.094									.432
2)	Entered	0.091	0.107								.593
3)	Entered	0.083	0.106	0.066							.706
4)	Entered	0.098	0.113	0.076	0.051						.757
5)	Entered	0.096	0.131	0.074	0.053	0.064					.801
6)	Entered	0.096	0.099	0.087	0.048	0.073	0.050				.845
7)	Entered	0.074	0.069	0.102	0.036	0.061	0.067	0.061			.876
8)	Entered	0.042	0.039	0.109	0.019	0.071	0.094	0.094	0.057		.919
9)	Removed	0.032	0.029	0.109		0.070	0.100	0.104	0.063		.914
10)	Removed	0.029		0.114		0.066	0.109	0.114	0.068		.909
11)	Removed			0.118		0.069	0.113	0.125	0.075		.902
12)	Entered			0.117		0.069	0.115	0.124	0.075	0.043	.934

ในการนำเสนอสารสนเทศโดยรวมด้วยการคัดเลือกแบบลำดับขั้น พบว่า ประกอบด้วย 12 รูปแบบ เริ่มจากการคัดเลือกตัวแปรอิสระที่มีสหสัมพันธ์กับตัวแปรตามสูงที่สุดเข้าสมการก่อน คือ “ความเหมาะสมของการเลือกใช้ภาพข้อมูลแต่ละประเภทในการนำเสนอ (X_{19})” ที่มีค่าสัมประสิทธิ์การถดถอย (B = .094) จากนั้นก็จะทดสอบตัวแปรที่ไม่ได้อยู่ในสมการ พบว่า มีตัวแปรทำนายที่เกี่ยวข้องในสมการอีก 8 ตัว รวมทั้งสิ้น 9 ตัว ได้แก่ X_{19}, X_{11}, X_5, X_{15}, X_{10}, X_4, X_{12}, X_7 และ X_2

จากตารางที่ 4 แสดงให้เห็นว่าตัวแปรอิสระที่มีสหสัมพันธ์กับตัวแปรตามสูงที่สุด คือ X_{19} ที่มีค่าสัมประสิทธิ์การถดถอยเท่ากับ .094 ร่วมกับตัวแปรที่ถูกคัดเลือกเข้ามาในสมการด้วยวิธีการคัดเลือกแบบก้าวหน้าอีก 7 ตัว ได้แก่ X_{11}, X_5, X_{15}, X_{10}, X_4, X_{12} และ X_7 มีค่าสัมประสิทธิ์สหสัมพันธ์ .919 และสามารถพยากรณ์ตัวแปรตามได้ร้อยละ 84.4 เมื่อใช้สถิติ F ทดสอบพบว่ามีนัยสำคัญทางสถิติ (sig F change < .05) แสดงว่าตัวแปรพยากรณ์ ทั้งหมดสามารถอธิบายความแปรปรวนของตัวแปรตามได้อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 อย่างไรก็ดี จะเห็นได้ว่า ตัวแปรที่อยู่ในสมการ 3 ตัวได้แก่ X_{15}, X_{11} และ X_{19} มีความผันผวน และมีแนวโน้มลดลงอย่างต่อเนื่องอันจะเห็นได้จากค่าสัมประสิทธิ์การถดถอยที่ลดลงเรื่อย ๆ ดังนั้นจึงทำการขจัดตัวแปรทั้ง 3 ออกจากสมการด้วยวิธีการคัดเลือกแบบถอยหลังที่ความน่าจะเป็น <= .050 และ >= .100 เมื่อพิจารณาค่า Partial Correlation ที่ใช้พิจารณาตัวแปรอิสระที่ถูกตัดออก มีค่าเท่ากับ .233, .215 และ .271 ตามลำดับ และทำการคัดเลือกตัวแปรเข้าสู่สมการอีกครั้ง คือ X_2 เพื่อสร้างแบบจำลองเชิงเส้นที่ดีที่สุดในการพยากรณ์ พบว่า มีค่าสัมประสิทธิ์สหสัมพันธ์เพิ่มขึ้นจากเดิมเป็น .934 และสามารถพยากรณ์ผลการทดสอบความสามารถในการนำเสนอสารสนเทศ ได้ร้อยละ 87.2 เมื่อใช้สถิติ F ทดสอบพบว่ามีนัยสำคัญทางสถิติ (sig F change < .05) แสดงว่าตัวแปรพยากรณ์ทั้งหมดสามารถอธิบายความแปรปรวนของตัวแปรตามอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ดังตารางที่ 5

ตารางที่ 5 แบบจำลองเชิงเส้นจากสมการวิเคราะห์การถดถอยพหุคูณแบบลำดับขั้น

ตัวแปร	B	SE_b	Beta	t	Sig.	Correlation		
						Zero-Order	Partial	Part
1) X_5	0.117	0.014	0.687	8.569	0.000	0.427	0.873	0.639
2) X_{10}	0.069	0.019	0.291	3.700	0.001	0.175	0.611	0.276
3) X_4	0.115	0.013	0.721	8.582	0.000	0.283	0.873	0.640
4) X_{12}	0.124	0.017	0.595	7.398	0.000	0.301	0.839	0.552
5) X_7	0.075	0.014	0.437	5.487	0.000	0.201	0.753	0.409
6) X_2	0.043	0.013	0.244	3.260	0.003	0.233	0.562	0.243
ค่าคงที่ = 1.795 ; SE_{est} = ±.075 ; **R = .934 ; R Square = .872 ; F = 26.140 ; p-value = .000 ;**								

จากตารางที่ 5 แบบจำลองเชิงเส้นจากการวิเคราะห์การถดถอยพหุคูณแบบลำดับขั้น ประกอบด้วยตัวแปรพยากรณ์เพียง 6 ตัวจากทั้งหมด 20 ตัว ได้แก่ X_5, X_{10}, X_4, X_{12}, X_7 และ X_2 จาก 3 องค์ประกอบหลัก ได้แก่ ด้านประสิทธิผล ด้านประสิทธิภาพ และด้านการเรียนรู้ ร่วมกันพยากรณ์ผลการทดสอบความสามารถในการนำเสนอสารสนเทศโดยรวมได้อย่างแม่นยำร้อยละ 87.2 และมีค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณเป็น .934 อย่างมีนัยสำคัญทางสถิติ และมีความคลาดเคลื่อนในการพยากรณ์ ±.075 เมื่อพิจารณาค่าน้ำหนักของตัวแปรอิสระ พบว่า ทั้งหมดเพิ่มสูงขึ้น โดยความถูกต้อง

ของ การแสดงผลข้อมูลตามเงื่อนไขการค้นหา (X_4) และความถูกต้องของผลการสกัดข้อมูลรายงานอุบัติเหตุ (X_5) จากองค์ประกอบด้านประสิทธิผลมีอิทธิพลต่อตัวแปรตามสูงที่สุด (Beta = 0.721 และ 0.687 ตามลำดับ)

เมื่อพิจารณาองค์ประกอบอื่น ๆ พบว่า ความรวดเร็วในการนำเสนอภาพข้อมูล (X_2) จากองค์ประกอบด้านประสิทธิภาพที่ถึงจะมีน้ำหนักที่ส่งผลต่อตัวแปรตามไม่มากนัก แต่มีค่าความเบี่ยงเบนออกจากสมการเส้นตรงน้อยที่สุด (SE_b = 0.013) จึงส่งผลให้มีโอกาสที่จะพบอิทธิพลของตัวแปรที่มีนัยสำคัญมากขึ้น เช่นเดียวกับตัวแปรความถูกต้องของ การจัดวางรายชื่อจังหวัดเพื่อการค้นหา (X_7) จากองค์ประกอบด้านประสิทธิผลที่ส่งผลต่อตัวแปรตามในระดับปานกลาง นอกจากนั้นตัวแปรความสามารถในการใช้งานระบบได้อย่างรวดเร็ว (X_{10}) และความสามารถในการทำความเข้าใจกับกระบวนการได้มาซึ่งผลลัพธ์ได้ด้วยตนเอง (X_{12}) จากองค์ประกอบด้านการเรียนรู้ที่แม้จะมีน้ำหนักที่ส่งผลต่อตัวแปรตามไม่มากนัก แต่ก็เป็นตัวแปรอิสระที่สามารถร่วมพยากรณ์หรือมีอิทธิพลต่อตัวแปรตามอย่างมีนัยสำคัญทางสถิติที่ .05 เช่นเดียวกับตัวแปรอื่นในสมการ

นอกจากนี้ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation) จากตารางที่ 5 ยังสามารถใช้ตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระ (Multicollinearity) โดยในส่วนของ Correlation จะให้ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson Correlation) ดังตารางที่ 5 นำเสนอค่า Zero-order พบว่าตัวแปรอิสระทุกตัวไม่มีความสัมพันธ์กันสูงเกินไป (>.80) แสดงว่าสามารถนำมาพยากรณ์ตัวแปรตามพร้อมกันได้ ส่วนค่าสัมประสิทธิ์สหสัมพันธ์บางส่วน (Partial Correlation) ใช้ประกอบการคัดเลือกตัวแปรอิสระในกรณีที่เกิดภาวะความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกัน โดยจะพิจารณาเลือกค่า Partial ที่มีค่าสูงกว่าไว้ จะเป็นวิธีที่เหมาะสมกว่าการนำตัวแปรอิสระเข้าสมการทั้งสองตัวพร้อมกัน นอกจากนั้น ยังอธิบายค่าสัมประสิทธิ์สหสัมพันธ์แบ่งส่วน (Part Correlation หรือ Semi-partial Correlation) ซึ่งหากนำมายกกำลังสองแล้วคูณด้วย 100 จะเป็นค่าที่บ่งบอกถึงร้อยละความแปรปรวนของตัวแปรเกณฑ์ที่อธิบายได้จากตัวแปรอิสระแต่ละตัว [44] แต่ค่าที่จะใช้ตรวจสอบภาวะความสัมพันธ์ของตัวแปรอิสระได้อย่างชัดเจนจริง ๆ ว่าแต่ละตัวนั้นไม่มีปัญหาเรื่องมีความสัมพันธ์กันสูงมากจนเกือบจะเป็นตัวเดียวกันกับค่า Tolerance และค่า VIF ในส่วนของตารางที่ 3

6. วิจารณ์ผลการวิจัย

ผลการวิเคราะห์สมการถดถอยพหุคูณแบบลำดับขั้นพบว่า สมการเชิงเส้นที่ดีที่สุด ประกอบด้วย 6 ตัวแปร ได้แก่ “ความถูกต้องของผลการสกัดข้อมูลรายงานอุบัติเหตุ”, “ความสามารถในการใช้งานระบบได้อย่างรวดเร็ว”, “ความถูกต้องของการแสดงผลข้อมูลตามเงื่อนไขการค้นหา”, “ความสามารถในการทำความเข้าใจกับกระบวนการได้มาซึ่งผลลัพธ์ได้ด้วยตนเอง”, “ความถูกต้องของการจัดวางรายชื่อจังหวัดเพื่อการค้นหา” และ “ความรวดเร็วในการนำเสนอภาพข้อมูล” ร่วมกันพยากรณ์ผลการทดสอบความสามารถในการนำเสนอสารสนเทศโดยรวม เมื่อพิจารณาลงไปในรายละเอียด พบว่า “ความถูกต้องของการแสดงผลข้อมูลตามเงื่อนไขการค้นหา” และ “ความถูกต้องของผลการสกัดข้อมูลรายงานอุบัติเหตุ” มีอิทธิพลต่อตัวแปรตามสูงที่สุด แสดงถึงผู้ใช้ให้ความสำคัญกับด้านประสิทธิผลของระบบเป็นหลัก ผลการทดสอบดังกล่าวสอดคล้องกับเป้าหมายการใช้งานที่กำหนดโดย ISO [20] ว่าหนึ่งในองค์ประกอบการใช้งานระบบเทคโนโลยีสารสนเทศที่เกี่ยวกับการทำงานของมนุษย์และคอมพิวเตอร์จะประกอบด้วยด้านประสิทธิผลที่ผู้ใช้บรรลุความถูกต้องของผลลัพธ์ในขอบเขตของระบบ โดยมุ่งความสนใจว่าผู้ใช้สามารถใช้งานระบบได้ถูกต้องสมบูรณ์ตามที่ได้ออกแบบไว้ภายใต้สถานการณ์ที่กำหนดได้หรือไม่ และเป็นไปในทิศทางเดียวกับเป้าหมายการใช้งานได้ที่กำหนดโดยนักวิจัย ได้แก่ Preece [22], Shneiderman และ Plaisant [23], Brinck และคณะ [24], Rubin และ Chisnell [25] รวมถึง Nielsen [26] ที่กำหนดองค์ประกอบด้านประสิทธิผลเป็นหนึ่งในเป้าหมายหลักของการใช้งานระบบร่วมกับองค์ประกอบด้านประสิทธิภาพ และด้านการเรียนรู้

สังเกตว่าตัวแปรอิสระจากองค์ประกอบด้านความยืดหยุ่น และด้านความพึงพอใจที่ถูกคัดเลือกเข้าสมการเริ่มต้น ได้แก่ “ความเหมาะสมของหน้าเว็บไซต์เมื่อแสดงผลบนเว็บเบราว์เซอร์ที่แตกต่างกัน (X_{15})” และโดยเฉพาะอย่างยิ่ง “ความเหมาะสมของการเลือกใช้ภาพข้อมูลแต่ละประเภท

ในการนำเสนอ (X_{19})" ที่ถูกคัดเลือกเข้าสมการเป็นตัวแรกจากค่าสัมประสิทธิ์การถดถอยที่มากที่สุด แต่มีค่าลดลงเรื่อย ๆ ในการปรับโมเดลแต่ละครั้ง จนถูกพิจารณาตัดออกจากสมการ ดังนั้นสมการที่ดีที่สุดในการพยากรณ์ผลการทดสอบความสามารถในการนำเสนอสารสนเทศโดยรวมมาจาก 3 องค์ประกอบหลัก ได้แก่ ด้านประสิทธิผล ด้านประสิทธิภาพ และด้านความสามารถในการเรียนรู้ สอดคล้องกับผลการวิจัยของ Folmer และ Bosch [45] ที่แยกองค์ประกอบความสามารถในการใช้งานตามบริบทของผู้ใช้เป็น 2 ส่วน ได้แก่ 1) ส่วนที่เป็นสมรรถนะของผู้ใช้ (Objective) เช่น ประสิทธิภาพของระบบ ประสิทธิผลที่ได้จากการใช้งาน และความสามารถในการเรียนรู้ของผู้ใช้งาน อันเกิดจากประสบการณ์การใช้งาน และ 2) ส่วนที่เป็นความคิดเห็นส่วนตัวของผู้ใช้ (Subjective) เช่น ความยืดหยุ่นในการใช้งาน และความพึงพอใจของผู้ใช้ที่มีลักษณะแบบปัจเจกบุคคล รายละเอียดดังตารางที่ 6

ตารางที่ 6 ผลการวิเคราะห์สมการถดถอยพหุคูณแบบลำดับขั้นเปรียบเทียบกับองค์ประกอบการใช้งาน

Usability Components		[20]	[21]	[22]	[19]	[23]	[24]	[25]	[26]	In this Article
สมรรถนะของผู้ใช้ [45]	Efficiency	✓		✓		✓	✓	✓	✓	✔
	Effectiveness	✓	✓	✓	✓	✓	✓	✓	✓	✔
	Learnability		✓	✓	✓	✓	✓	✓	✓	✔
ความคิดเห็นส่วนตัวของผู้ใช้	Flexibility		✓		✓					
	Satisfaction	✓	✓			✓	✓		✓	

จากตารางที่ 6 ผลวิเคราะห์การถดถอยพหุคูณแบบลำดับขั้นจากผู้ใช้แสดงให้เห็นว่าองค์ประกอบการใช้งานที่ส่งผลต่อผลการทดสอบความสามารถในการนำเสนอสารสนเทศโดยรวม มี 3 องค์ประกอบ และทั้งหมดเป็นส่วนที่เป็นสมรรถนะของผู้ใช้ทั้งสิ้น แสดงถึงผู้ใช้ส่วนใหญ่มีความคิดเห็นไปในทิศทางเดียวกัน ซึ่งตรงกันข้ามกับส่วนที่เป็นความคิดเห็นส่วนตัวของผู้ใช้ที่มีความคิดเห็นแตกต่างกัน ผลการทดสอบดังกล่าวสอดคล้องกับบริบทของผู้ใช้ที่แตกต่างกัน สามารถแบ่งได้ 3 กลุ่ม คือ กลุ่มผู้เชี่ยวชาญด้านคอมพิวเตอร์ที่มีประสบการณ์ใช้งานคอมพิวเตอร์มากกว่า 15 ปี กลุ่มผู้ทำหน้าที่ข้อมูลจราจร ที่มีประสบการณ์ในการใช้งานระบบด้านการจราจรที่เกี่ยวข้องมากกว่า 15 ปี และกลุ่มผู้ใช้งานทั่วไป

7. สรุปและวิจารณ์ผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อสกัดข้อมูลรายงานอุบัติเหตุทางถนน และทดสอบความสามารถในการนำเสนอสารสนเทศที่เกี่ยวข้อง การดำเนินงานเริ่มจากการรวบรวมข้อมูลรายงานอุบัติเหตุทางถนนรายใหญ่จากเว็บไซต์ เข้าสู่การเตรียมข้อมูลเพื่อการวิเคราะห์ข้อความ การจัดการกับคำที่เขียนไม่ถูกต้องด้วยการวัดความคล้ายคลึงของคำด้วยเทคนิคการวัดระยะทางเลเวนชเตย์น เพื่อสร้างคลังข้อมูลคำศัพท์ร่วมกับการรู้จำเอนทิตี้มาสกัดข้อมูลใน 6 ประเด็น ได้แก่ จังหวัด วันที่ จำนวนผู้เสียชีวิต จำนวนผู้บาดเจ็บ ประเภทและจำนวนของยานพาหนะที่เกิดเหตุ เพื่อนำเสนอสารสนเทศด้วยภาพข้อมูลปฏิสัมพันธ์กับผู้ใช้ผ่านทางเว็บไซต์ เมื่อทดสอบความสามารถในการนำเสนอสารสนเทศรายงานอุบัติเหตุทางถนนด้านประสิทธิผลของการค้นคืนข้อมูล พบว่า ค่าความถูกต้อง ค่าความครบถ้วน และค่าประสิทธิภาพโดยรวม เท่ากับ 0.87, 0.85 และ 0.86 ตามลำดับ

ผลการทดสอบความสามารถในการนำเสนอสารสนเทศเลือกใช้วิธีการสุ่มตัวอย่างโดยไม่ใช้ความน่าจะเป็นจากการเลือกกลุ่มตัวอย่างแบบเฉพาะเจาะจงด้วยการเลือกสรร

ผู้เชี่ยวชาญตามบริบทในการใช้งานสารสนเทศการสกัดข้อมูลรายงานอุบัติเหตุทางถนนออกเป็น 3 กลุ่ม กลุ่มละ 10 คน ได้แก่ กลุ่มผู้เชี่ยวชาญด้านคอมพิวเตอร์ กลุ่มผู้ทำหน้าที่ข้อมูลจราจร และกลุ่มผู้ใช้งานทั่วไป รวมทั้งสิ้น 30 คน ผลการทดสอบความสามารถในการนำเสนอสารสนเทศโดยรวมอยู่ในระดับดี ก่อนการวิเคราะห์ข้อมูล ได้มีการตรวจสอบข้อตกลงเบื้องต้นในการวิเคราะห์การถดถอย เมื่อวิเคราะห์ถดถอยพหุคูณแบบลำดับขั้น พบว่า ประกอบด้วย 6 ตัวแปร ได้แก่ “ความถูกต้องของผลการสกัดข้อมูลรายงานอุบัติเหตุ”, “ความสามารถในการใช้งานระบบได้อย่างรวดเร็ว”, “ความถูกต้องของการแสดงผลข้อมูลตามเงื่อนไขการค้นหา”, “ความสามารถในการทำความเข้าใจกับกระบวนการได้มาซึ่งผลลัพธ์ได้ด้วยตนเอง”, “ความถูกต้องของการจัดวางรายชื่อจังหวัดเพื่อการค้นหา” และ “ความรวดเร็วในการนำเสนอภาพข้อมูล” ร่วมกันพยากรณ์ผลการทดสอบความสามารถในการนำเสนอสารสนเทศโดยรวม เมื่อพิจารณาในรายละเอียด พบว่า “ความถูกต้องของการแสดงผลข้อมูลตามเงื่อนไขการค้นหา” และ “ความถูกต้องของผลการสกัดข้อมูลรายงานอุบัติเหตุ” มีอิทธิพลต่อตัวแปรตามสูงที่สุด แสดงถึงผู้ใช้ให้ความสำคัญกับด้านประสิทธิผลของระบบเป็นหลัก ร่วมกับองค์ประกอบด้านประสิทธิภาพ และด้านการเรียนรู้ ในทางกลับกัน พบว่า ไม่มีตัวแปรจากองค์ประกอบด้านความยืดหยุ่น และด้านความพึงพอใจที่ถูกคัดเลือกเข้าสมการพยากรณ์ แสดงให้เห็นถึงผู้ใช้ส่วนใหญ่มีความคิดเห็นไปในทิศทางเดียวกันในส่วนที่เป็นสมรรถนะของผู้ใช้ เช่น ประสิทธิภาพ ประสิทธิผล และการเรียนรู้ ซึ่งตรงกันข้ามกับส่วนที่เป็นความคิดเห็นของผู้ใช้ที่มีความแตกต่างกันตามบริบท

8. ข้อเสนอแนะ

ผลลัพธ์จากการสกัดข้อมูลรายงานอุบัติเหตุรายใหญ่ในงานวิจัยนี้ ไม่สามารถตรวจสอบความถูกต้องด้วยการเปรียบเทียบกับสถิติการเกิดอุบัติเหตุจากแหล่งอ้างอิงของหน่วยงานใดได้ เนื่องมาจากข้อมูลที่ปรากฏบนเว็บไซต์ของบริษัทกลางคุ้มครองผู้ประสบภัยจากรถ จำกัด นั้นเก็บข้อมูลเฉพาะผู้ที่มาใช้สิทธิตามพระราชบัญญัติคุ้มครองผู้ประสบภัยจากรถเท่านั้น อย่างไรก็ดี งานวิจัยนี้มุ่งนำเสนออีกแนวทางในการสกัดข้อมูลกึ่งมีโครงสร้างที่สามารถนำไปประยุกต์ใช้กับข้อมูลลักษณะเดียวกัน ข้อค้นพบจากงานวิจัยนี้มุ่งนำเสนอคุณภาพของการนำเสนอสารสนเทศตามที่ผู้ใช้ต้องการ ซึ่งสามารถนำไปประยุกต์ในการนำเสนอสารสนเทศด้วยภาพข้อมูล เพื่อให้เหมาะสมและตรงกับความต้องการของผู้ใช้มากยิ่งขึ้น สิ่งที่ควรพัฒนาต่อไปคือการนำออนโทโลยีเข้ามาช่วยอำนวยความสะดวกแก่ผู้ใช้ในการค้นคืนสารสนเทศด้วยหลักการค้นหาเชิงความหมาย ซึ่งจะดำเนินการต่อไปในอนาคต

9. เอกสารอ้างอิง

1. World Health Organization, 2018, Global Status Report on Road Safety 2018, World Health Organization, Geneva, p. 245

2. Division of Non Communicable Disease, 2015, Where did the Data and Statistics of the Dead Disappear?, Bureau of Non Communicable Disease, Nonthaburi, pp. 1-6. (In Thai)

3. Road Safety Center, 2017, Accident data definition, Department of Disaster Prevention and Mitigation, Ministry of Interior, Bangkok, pp. 1-6. (In Thai)

4. Devadason, F.G., 1996, “Practical Steps for Identifying Information Needs of Clients”, *Proceedings of the Tenth Congress of Southeast Asian Librarians (CONSAL X)*, Kuala Lumpur, Malaysia.

5. Ware, C., 2020, Information Visualization: Perception for Design, 4th ed., Morgan Kaufmann, Boston.

6. McConnell, S., 1996, Rapid Development: Taming Wild Software Schedules, Microsoft Press, Washington.

7. Mitchell, R., 2015, Web Scraping with Python. 2nd ed., O’Reilly Media, Sebastopol.

8. Garson, G.D., 2015, Missing Values Analysis and Data Imputation, Statistical Associates Publishing, Asheboro.

9. Tapsai, C., Unger, H. and Meesad, P., 2021, Thai Natural Language Processing Word Segmentation, Semantic Analysis, and Application, Springer, Cham.

10. Porter, M.F., 1980, "An Algorithm for Suffix Stripping," *Program: Electronic Library and Information Systems*, 14 (3), pp. 130-137.

11. Santirattanaphakdi, C., 2020, "The Design and Development of a Knowledge Extraction and Retrieval System via Data Visualization Case Study: Road Major Accidents on Website," *APHEIT Journal Science and Technology*, 9 (2), pp. 17-32. (In Thai)

12. Levenshtein, V. I., 1966, "Binary Codes Capable of Correcting Deletions, Insertions, and Reversals," *Soviet Physics Doklady*, 10 (8), pp. 707–0710.

13. Miller, F.P., Vandome, A.F. and McBrewster, J., 2009, Levenshtein Distance, Alphascript Publishing, Gladbach.

14. Indurkhya, N. and Damerau, F.J., 2010, Handbook of Natural Language Processing, 2nd ed., Chapman & Hall, New York.

15. Martinez, J., 2018, Google Charts for Institutional Research Websites, University of Houston, Houston.

16. Chen, C., Härdle, W. and Unwin, A., 2008, Handbook of Data Visualization, Springer, Berlin.

17. Wilke, C.O., 2019, Fundamentals of Data Visualization A Primer on Making Informative and Compelling Figures, O'Reilly Media, Boston.

18. Manning, C.D., Raghavan, P. and Schütze, H., 2008, Introduction to Information Retrieval, 1st ed., Cambridge University Press, Cambridge.

19. Dix, A., Finlay, J., Abowd, G.D. and Beale, R., 2004, Human–computer Interaction, 3rd ed., Pearson Education Limited, Haddington.

20. The International Organization for Standardization (ISO), 1998, Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) Part 11: Guidance on Usability. ISO 1998, Genève.

21. Shackel, B. and Richardson, S., 1991, Human Factors for Informatics Usability – Background and Overview, Cambridge University Press, Cambridge.

22. Preece, J., 2002, Interaction Design beyond Human-computer Interaction, John Wiley & Sons, New York.

23. Shneiderman, B. and Plaisant, C., 2005, Designing the User Interface: Strategies for Effective Human-computer Interaction, 4th ed., Addison Wesley Longman, Massachusetts.

24. Brinck, T., Gergle, D. and Wood, S.D., 2012, Usability for the Web: Designing Web Sites that Work, Morgan Kaufmann, San Francisco.

25. Rubin, J. and Chisnell, D., 2008, Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests, Wiley, New York.

26. Nielsen, J., 1993, Usability Engineering, Morgan Kaufmann, New York.

27. Srisa-ard, O., 1995, "Validation of Tool Measurement by Expert," *Journal of Educational Measurement Mahasarakham University*, 1 (1), pp. 45-49. (In Thai)

28. Pipitkun, K., 2018, "Quality of Questionnaire Tools: Validity and Reliability in Public Administration for Research," *NEU Academic and Research Journal*, 8 (2), pp. 104-110. (In Thai)

29. Cronbach, L.J., 1990, Essentials of Psychological Testing, 5th ed., Harper Collins Publishers, New York.

30. Kanjanawasee, S., 2001, Selection of Appropriate Statistics for Research, Boonsiri Printing, Bangkok. (In Thai)

31. Nielsen, J. and Landauer, T.K, 1993, "A Mathematical Model of the Finding of Usability Problems," *The Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, Amsterdam, The Netherlands, pp. 206-213.

32. Saohin, W., 2001, "Sampling Technique and Sample Size", *Suddhiparitad*, 15 (47), pp. 80-87. (In Thai)

33. Jensen, C., 1996, Delphi in Depth: Power Techniques from the Experts, McGraw-Hill, Berkeley.

34. Jensantikul, N., 2017, "Utilizing the Delphi Technique for Research", *Kasetsart University Political Science Review Journal*, 4 (2), pp. 48-64. (In Thai)

35. Likert, R., 1961, New Patterns of Management, McGraw-Hill Book Company, New York.

36. Srisa–ard, B., 2011, Basic Research, Suveiriyasan Publishing, Bangkok. (In Thai)

37. Ghorbani, H., 2019, "Mahalanobis Distance and Its Application for Detecting Multivariate Outliers," *Series: Mathematics and Informatics*, 34 (3), pp. 583-595.

38. Steinskog, D. J., Tjøstheim, D. B., and Kvamstø, N.G., 2007, "A Cautionary Note on the Use of the Kolmogorov–Smirnov Test for Normality," *American Meteorological Society*, 135 (3), pp. 1151–1157.

39. Islam, M.R., 2018, "Sample Size and Its Role in Central Limit Theorem (CLT)," *Computational and Applied Mathematics Journal*, 4 (1), pp. 1-7.

40. Razali, N.M. and Wah, Y.P., 2011, "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests," *Journal of Statistical Modeling and Analytics*, 2 (1), pp. 21-33.

41. Tabachnick, B.G. and Fidell, L.S., 2001, Using Multivariate Statistics, Allyn & Bacon, Needham.

42. Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E., 2014, Multivariate Data Analysis, 7th ed., Pearson Education, Harlow.

43. Wiratchai, N., 2010, Course 21701 Curriculum and Instruction Research Unit 7 Studies of Relevant Literature and Unit 10 Quantitative Analytical Statistics: Descriptive Statistics and Parametric Statistics Master of Education Degree Program Courses and Teaching Sukhothai Thammathirat University, Sukhothai Thammathirat Publishing, Bangkok. (In Thai)

44. Phusee-orn, S., 2018, Applying SPSS to Analyze Research Data, 8th ed., Taksila Publishing, Mahasarakham. (In Thai)

45. Folmer, E. and Bosch, J., 2004, "Architecting for Usability: a Survey," *The Journal of Systems and Software*, 70 (1-2), pp. 61-78.