



การพัฒนากรอบการทำงานหาความสัมพันธ์คำถามคำตอบอัตโนมัติสำหรับห้องเรียนชุมชน

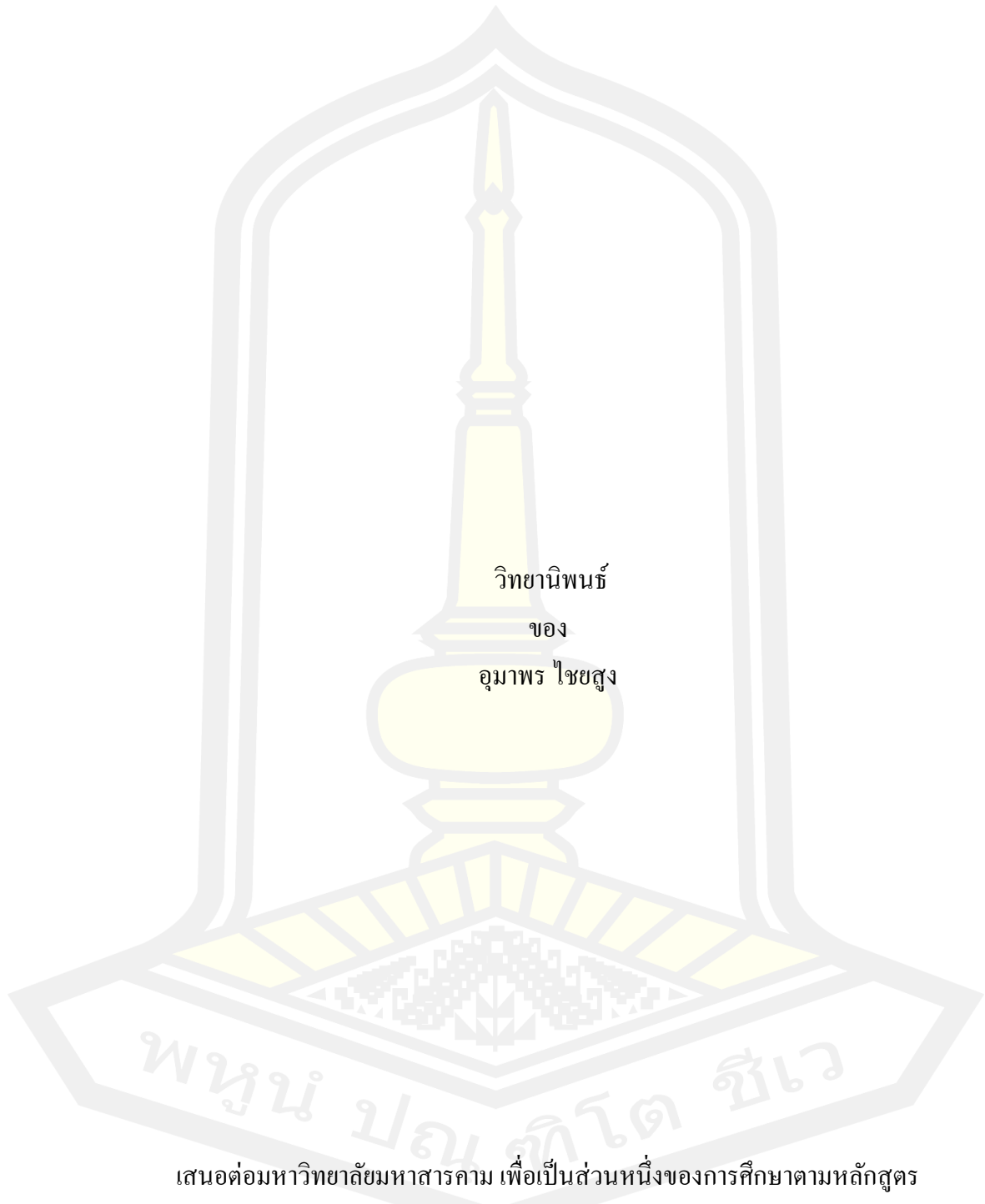
วิทยานิพนธ์
ของ
อุมพร ไชยสูง

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

พฤษภาคม 2565

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

การพัฒนากรอบการทำงานหาความสัมพันธ์คำถามคำตอบอัตโนมัติสำหรับห้องเรียนชุมชน



เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

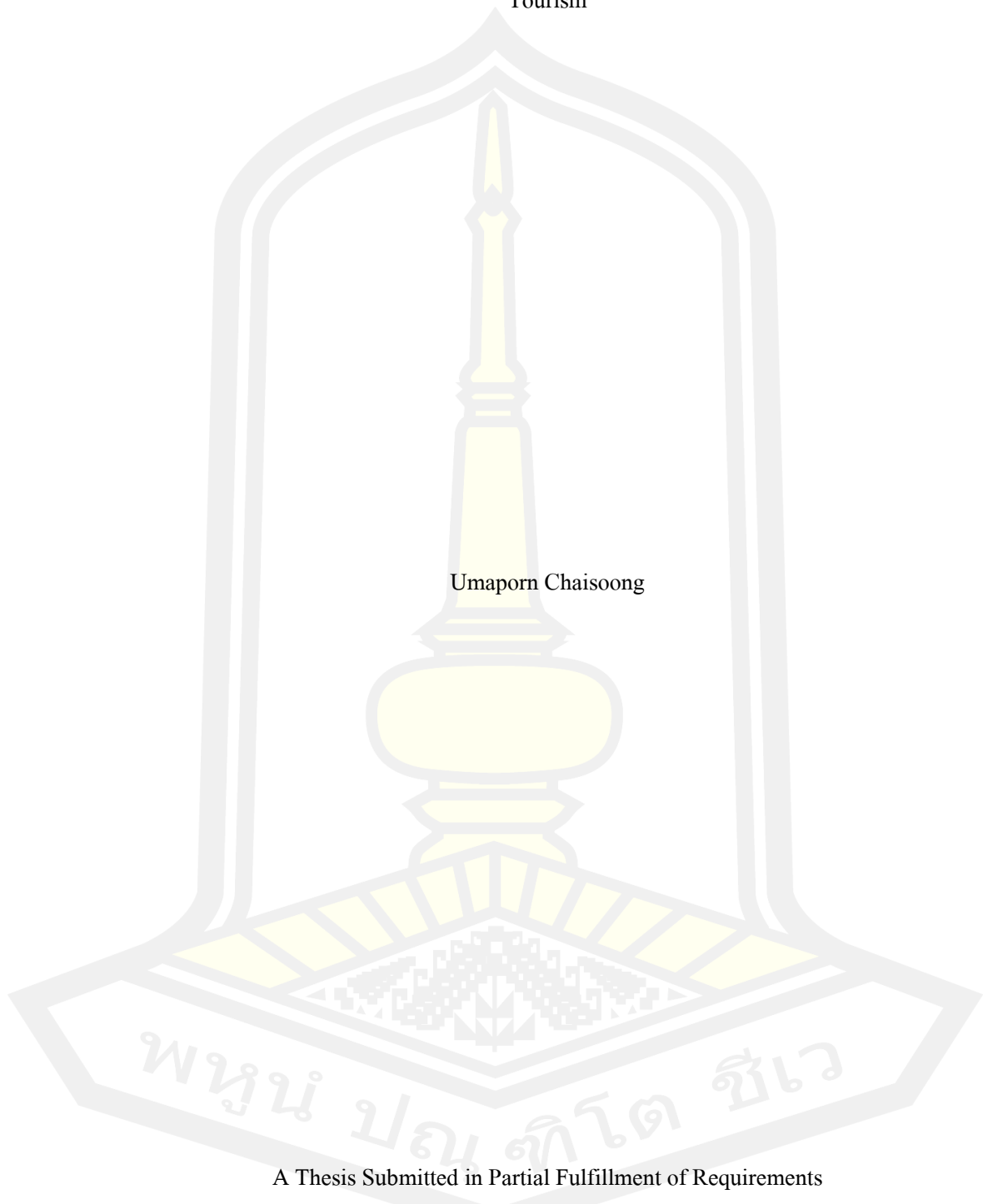
พฤศจิกายน 2565

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

Developing an Association Framework of Automatic Question and Answering for Community

Tourism

Umaporn Chaisoong



A Thesis Submitted in Partial Fulfillment of Requirements
for Doctor of Philosophy (Information Technology)

November 2022

Copyright of Mahasarakham University



คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนางสาวอุมาพร ไชยสูง
แล้วเห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชา
เทคโนโลยีสารสนเทศ ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

ประธานกรรมการ

(รศ. ดร. วรรัตน์ สงฆ์แป้น)

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผศ. ดร. นัตถเกล้า เจริญผล)

กรรมการ

(ผศ. ดร. โอพาริก สุรินตะ)

กรรมการ

(ผศ. ดร. พัฒนพงษ์ ชมภูวิเศษ)

กรรมการ

(ผศ. ดร. แกมกาญจน์ สมประเสริฐศรี)

มหาวิทยาลัยอนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตาม
หลักสูตรปริญญา ปรัชญาดุษฎีบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ ของมหาวิทยาลัย
มหาสารคาม

(ผศ. ศศิธร แก้วมัน)

คณบดีคณะวิทยาการสารสนเทศ

(รศ. ดร. กริสน์ ชัยมุล)

คณบดีบัณฑิตวิทยาลัย

ชื่อเรื่อง	การพัฒนากรอบการทำงานหาความสัมพันธ์คำถามคำตอบอัตโนมัติสำหรับ ท่องเที่ยวชุมชน		
ผู้วิจัย	อุมพร ไชยสูง		
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. นัตริเกล้า เจริญผล		
ปริญญา	ปรัชญาคุษฎีบัณฑิต	สาขาวิชา	เทคโนโลยีสารสนเทศ
มหาวิทยาลัย	มหาวิทยาลัยมหาสารคาม	ปีที่พิมพ์	2565

บทคัดย่อ

การค้นคืนสารสนเทศ (Information Retrieval: IR) เป็นแขนงหนึ่งของการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) วัตถุประสงค์ของการพัฒนากรอบการทำงานหาความสัมพันธ์คำถามคำตอบอัตโนมัติสำหรับท่องเที่ยวชุมชน เพื่อแก้ไขปัญหาการเข้าถึงข้อมูลเชิงพื้นที่ของนักท่องเที่ยวโดยชุมชนอีสานใต้ ถือเป็นความท้าทายของการประมวลผลภาษาไทยและปัญหาของการเขียนภาษาไทย อาทิเช่น การเขียนติดกัน โดยไม่มีช่องว่างหรือการแบ่งแยกคำ (Tokenize) ดังนั้นการตัดแบ่งคำที่ถูกต้องจึงส่งผลต่อความแม่นยำและความถูกต้องในการเข้าถึงข้อมูลอย่างมีประสิทธิภาพของนักท่องเที่ยว งานวิจัยในครั้งนี้จึงนำวิธีการวัดความคล้ายคลึง (Cosine Similarity) ทำงานบนฐานของ Vector Space Model (VSM) ที่ใช้กันอย่างแพร่หลาย โดยนำจุดเด่นของ Text Vectorization มาคำนวณเพื่อให้ได้มาซึ่งคุณสมบัติของคำสามารถเป็นตัวแทนเอกสารอย่างมีประสิทธิภาพ ผลการทดลองพบว่า ได้ถูกคำศัพท์ (Bag of Words: BoW) จำนวน 19,501 คำ (Term) จากเอกสารทั้งหมด 1,237 เอกสาร สำหรับการประเมินประสิทธิภาพของแบบจำลองให้ค่าความถูกต้อง (Accuracy) ร้อยละ 99 บวกถึงความสามารถในการอธิบายคำตอบและประสิทธิภาพของแบบจำลองได้อย่างแม่นยำ

คำสำคัญ : การค้นคืนเอกสาร, การตัดคำ, TF-IDF, การวัดความคล้ายคลึงแบบโคไซน์, การประมวลผลภาษาธรรมชาติ, การท่องเที่ยว

TITLE	Developing an Association Framework of Automatic Question and Answering for Community Tourism		
AUTHOR	Umaporn Chaisoong		
ADVISORS	Assistant Professor Chatklaw Jareanpon , Ph.D.		
DEGREE	Doctor of Philosophy	MAJOR	Information Technology
UNIVERSITY	Maharakham University	YEAR	2022

ABSTRACT

Information retrieval (IR) is one of the Natural Language Processing (NLP). The purpose of the research is to develop an association framework of automatic questions and answers resolving the problem of spatial information access of tourists in a case study of the South Northeastern Region, THAILAND. The difficulty of Thai language processing is that the language has no space in a sentence, without punctuation or word separation. Correctly tokenizing or separating words affects precision and accuracy. The efficient data access of the tourists is a challenge of this research. The similarity method called the Cosine Similarity technique is based on the Vector Space Model (VSM) well-known and efficient, by bringing the outstanding points of Text Vectorization to calculate and acquire the crucial features for being the document representatives efficiently. The result in this stage is Bag of Words, in all 19,501 terms, from 1,237 documents. For the evaluation of model effectiveness an Accuracy value of 96%, which best indicates the ability to describe the answer and effectiveness of the model.

Keyword : Information retrieval, Thai word segmentation, TF-IDF, Cosine similarity, Natural language processing, Tourism

กิตติกรรมประกาศ

งานวิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาระดับปริญญาเอก หลักสูตรปรัชญาดุษฎีบัณฑิตสาขาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม การดำเนินการวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดีและเป็นไปตามวัตถุประสงค์ของหลักสูตร ด้วยความอนุเคราะห์หืออย่างสูงยิ่งจากที่ปรึกษาหลัก ผู้ช่วยศาสตราจารย์ ดร.ฉัตรเกล้า เจริญผล และคณะกรรมการผู้ทรงคุณวุฒิภายใน ผู้ช่วยศาสตราจารย์ ดร.โอพาริก สุรินตะ ผู้ช่วยศาสตราจารย์ ดร.พัฒนพงษ์ ชมพูวิเศษ ผู้ช่วยศาสตราจารย์ ดร.แกมกาญจน์ สมประเสริฐศรี และผู้ทรงคุณวุฒิภายนอก รองศาสตราจารย์ ดร. วรรัตน์ สงฆ์แป้น ประธานกรรมการ ที่ได้แนะนำและให้คำปรึกษาแนวทางที่เป็นประโยชน์อย่างยิ่งต่อการแก้ไขข้อบกพร่องต่าง ๆ ด้วยความเอาใจใส่เป็นอย่างดีตลอดมาตั้งแต่ต้นจนสำเร็จเรียบร้อย รวมถึงมหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน ดันสังกัดที่ให้โอกาสในการพัฒนาศักยภาพตนเอง ผู้วิจัยขอกราบขอบพระคุณไว้เป็นอย่างสูง คุณค่าและประโยชน์จากวิทยานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบเป็นเครื่องบูชาคุณบิดา มารดา ที่เคารพยิ่ง ตลอดจนบูรพาจารย์และผู้มีพระคุณที่ให้การอบรมสั่งสอนให้ผู้วิจัยประสบความสำเร็จในปัจจุบันและเจริญก้าวหน้าในอนาคตต่อไป

อุมาพร ไชยสูง

พหุ น บณุ ทิโต ชีเว

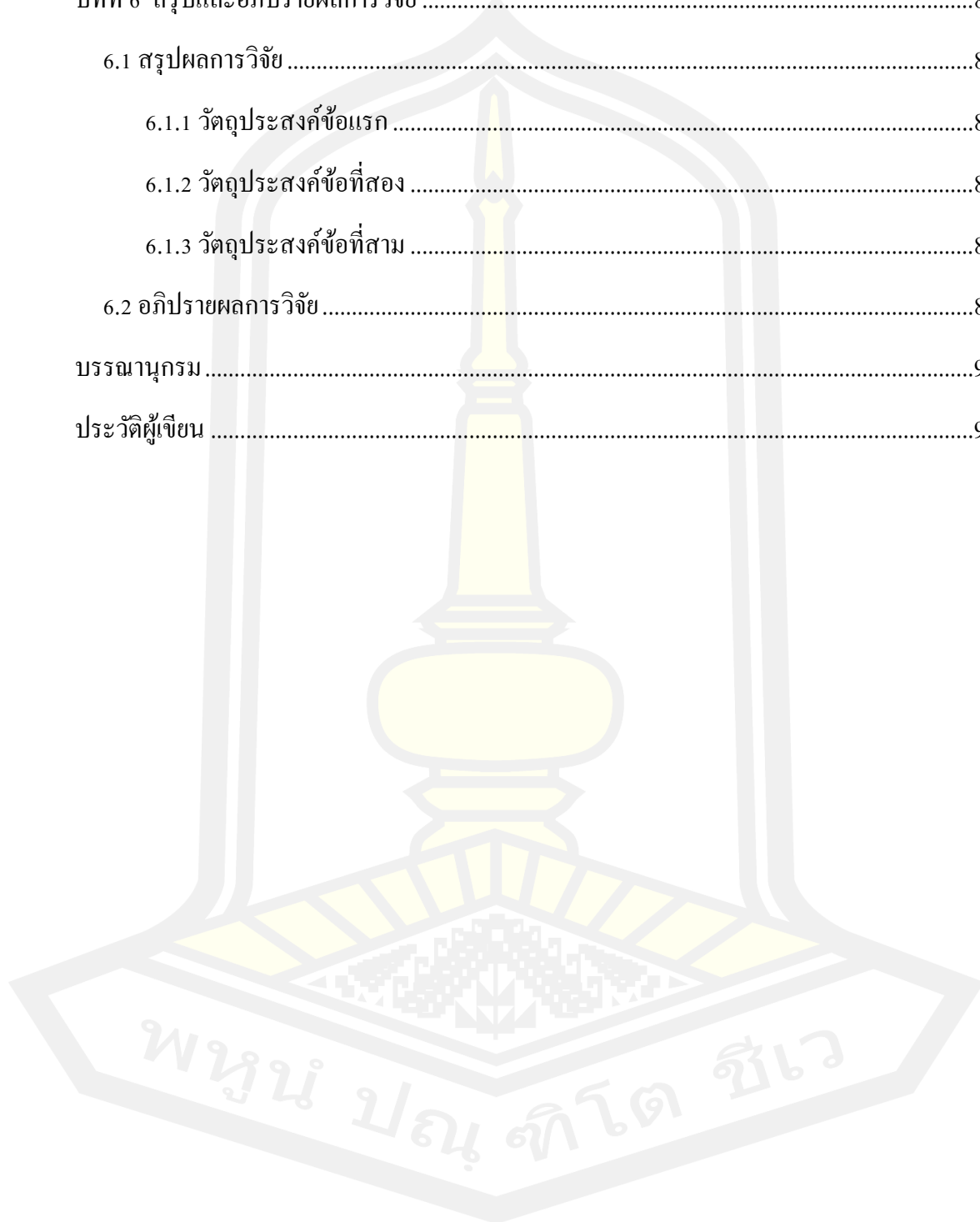
สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญ	ช
สารบัญตาราง	ฅ
สารบัญรูปภาพ	ฉ
บทที่ 1 บทนำ	1
1.1 ที่มาและความสำคัญของปัญหา	1
1.2 โจทย์วิจัย (Research question).....	7
1.3 จุดมุ่งหมายของการวิจัย	8
1.4 ขอบเขตการวิจัย.....	8
1.5 วิธีการดำเนินวิจัย	8
1.6 ผลที่คาดว่าจะได้รับ	9
1.7 นิยามศัพท์.....	9
1.8 รายละเอียดของวิทยานิพนธ์	10
บทที่ 2 วิธีการดำเนินงานวิจัย	12
2.1 การวิเคราะห์องค์ประกอบของกลุ่มรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดย ชุมชนอีสานใต้.....	13
2.1.1 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	13
2.1.2 การกำหนดขอบเขตปัญหาของงานวิจัย	14
2.1.3 การเก็บรวบรวมข้อมูล.....	14

2.1.4	วิธีการวิเคราะห์ข้อมูล.....	16
2.2	การค้นหารูปแบบคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้.....	17
2.2.1	แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	17
2.2.2	การกำหนดขอบเขตปัญหาของงานวิจัย.....	17
2.2.3	การเก็บรวบรวมข้อมูล (Data collection).....	18
2.2.4	วิธีการวิเคราะห์ข้อมูลพอสังเขป.....	21
2.3	การพัฒนากรอบการทำงานการหาคำตอบสำหรับตอบคำถามอัตโนมัติการท่องเที่ยวโดยชุมชนอีสานใต้.....	21
2.3.1	แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	21
2.3.2	การกำหนดขอบเขตปัญหาของงานวิจัย.....	22
2.3.3	การเก็บรวบรวมข้อมูล.....	23
2.3.4	วิธีการวิเคราะห์ข้อมูลพอสังเขป.....	23
บทที่ 3	การวิเคราะห์ห้วงค์ประกอบกลุ่มรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้.....	26
3.1	ที่มาและความสำคัญของปัญหา.....	26
3.2	แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	27
3.2.1	แนวคิดและทฤษฎีด้านการท่องเที่ยว (Tourism concept).....	27
3.2.2	การวิเคราะห์ปัจจัย (Factor analysis: FA).....	34
3.3	ระเบียบวิธีวิจัย.....	36
3.3.1	คำถามวิจัย (Question research).....	36
3.3.2	วัตถุประสงค์การวิจัย (Objective).....	37
3.3.3	กรอบแนวคิดการวิจัย (Research framework).....	37
3.3.4	ขอบเขตด้านประชากรและกลุ่มตัวอย่าง.....	37
3.3.5	เครื่องมือที่ใช้ในการวิจัย.....	38

3.4 ผลการทดลอง	41
3.5 สรุปผลการทดลอง.....	43
บทที่ 4 การค้นหารูปแบบคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้	47
4.1 ที่มาและความสำคัญของปัญหา	47
4.2 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	48
4.2.1 เหมืองข้อมูล (Data Mining: DM).....	48
4.2.2 เหมืองข้อความ (Text Mining: TM).....	49
4.2.3 กฎความสัมพันธ์ (Association Rule Mining: ARM).....	49
4.2.4 เอพริออริ (Apriori algorithm).....	50
4.3 ระเบียบวิธีวิจัย	50
4.3.1 คำถามวิจัย (Question research)	50
4.3.2 วัตถุประสงค์การวิจัย (Objective).....	50
4.3.3 กรอบแนวคิดการวิจัย (Research framework).....	50
4.3.4 การเก็บรวบรวมข้อมูล (Data collection).....	51
4.3.5 การเตรียมข้อมูล (Data preparation)	52
4.4 ผลการทดลอง	52
4.5 สรุปผลการทดลอง.....	55
บทที่ 5 การพัฒนากรอบการทำงานการหาคำตอบสำหรับตอบคำถามอัตโนมัติการท่องเที่ยว โดยชุมชนอีสานใต้.....	58
5.1 ที่มาและความสำคัญของปัญหา	58
5.2 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	59
5.2.1 แนวคิดและทฤษฎีที่เกี่ยวข้อง	59
5.2.2 งานวิจัยที่เกี่ยวข้อง.....	65
5.3 ระเบียบวิธีวิจัย	67

5.4 ผลการทดลอง	72
บทที่ 6 สรุปและอภิปรายผลการวิจัย	80
6.1 สรุปผลการวิจัย	80
6.1.1 วัตถุประสงค์ข้อแรก	80
6.1.2 วัตถุประสงค์ข้อที่สอง	81
6.1.3 วัตถุประสงค์ข้อที่สาม	82
6.2 อภิปรายผลการวิจัย	83
บรรณานุกรม	96
ประวัติผู้เขียน	98



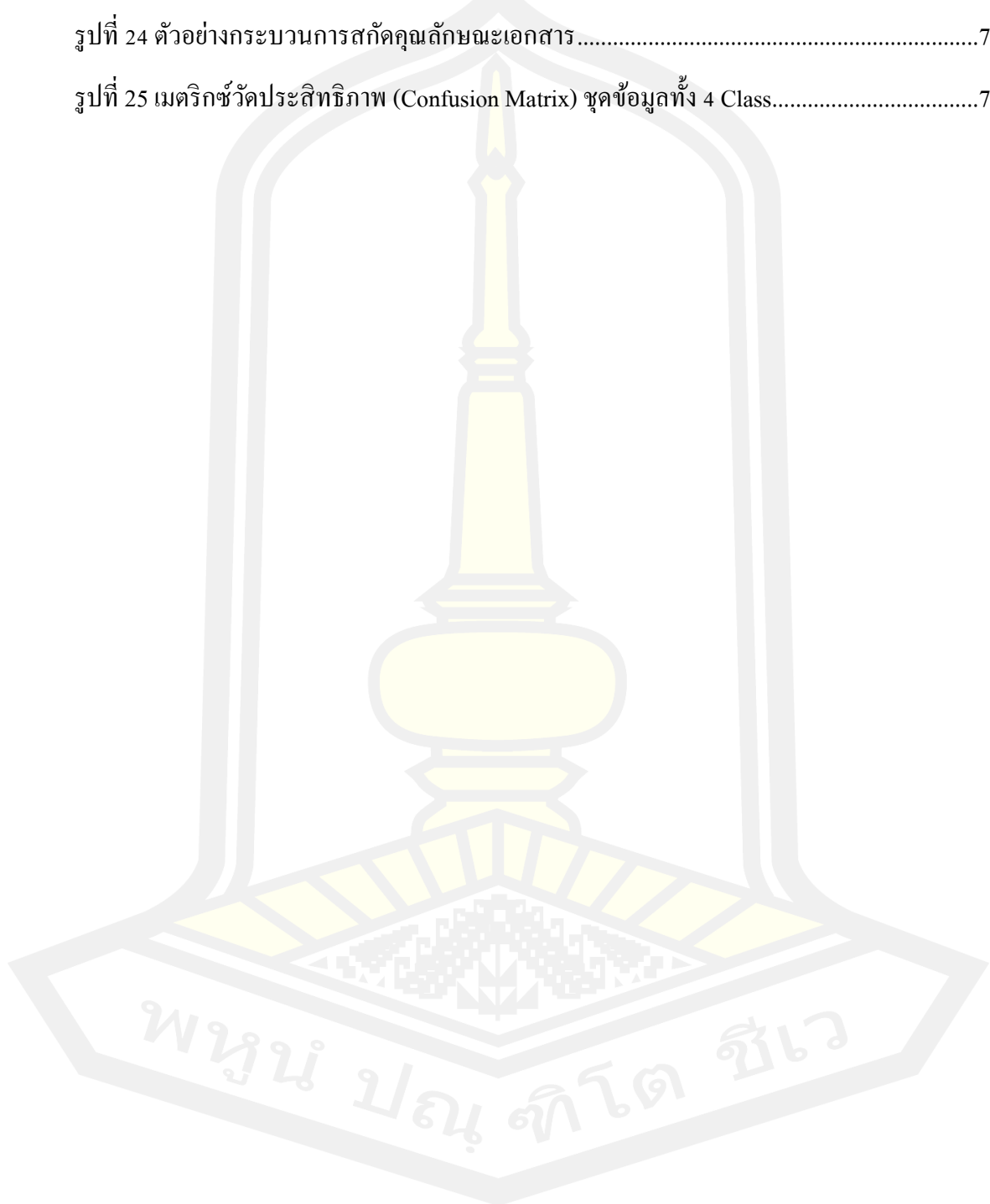
สารบัญตาราง

	หน้า
ตารางที่ 1 รายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวอิสานใต้ 50 รายการคำถาม	19
ตารางที่ 2 การแทนค่าข้อมูลน้ำหนักระแนงด้วยเลข 0 และ 1 (การเตรียมข้อมูล)	20
ตารางที่ 3 ข้อมูลการตอบแบบสอบถามที่ถูกแทนค่าของทั้ง 4 กลุ่ม (การเตรียมข้อมูล)	20
ตารางที่ 4 รูปแบบรายการคำถามที่มีความสัมพันธ์กับการท่องเที่ยวอิสานใต้ Class E	56
ตารางที่ 5 รูปแบบรายการคำถามที่มีความสัมพันธ์กับการท่องเที่ยวอิสานใต้ Class S	56
ตารางที่ 6 รูปแบบรายการคำถามที่มีความสัมพันธ์กับการท่องเที่ยวอิสานใต้ Class A	56
ตารางที่ 7 รูปแบบรายการคำถามที่มีความสัมพันธ์กับการท่องเที่ยวอิสานใต้ Class N	57
ตารางที่ 8 วิเคราะห์จุดเด่น-ด้อย วิธีการค้นคืนสารสนเทศ	60
ตารางที่ 9 เปรียบเทียบวิธีการค้นคืนสารสนเทศและการประเมินประสิทธิภาพ	65
ตารางที่ 10 แสดงรายการชุดเอกสารคำตอบ (Answer collections) ทั้ง 4 กลุ่ม (Cluster)	71
ตารางที่ 11 แสดงตัวอย่างเวกเตอร์คำน้ำหนักรำ (Term vector) ทั้ง 4 กลุ่ม (Cluster)	74
ตารางที่ 12 แสดงค่าความคล้ายคลึง 3 ลำดับแรก กลุ่ม E ด้านการบริการที่เป็นเลิศ	75
ตารางที่ 13 แสดงค่าความคล้ายคลึง 3 ลำดับแรก กลุ่ม S ด้านสิ่งอำนวยความสะดวกที่ได้มาตรฐาน	75
ตารางที่ 14 แสดงค่าความคล้ายคลึง 3 ลำดับแรก กลุ่ม A ด้านข้อมูลสารสนเทศและประชาสัมพันธ์	76
ตารางที่ 15 แสดงค่าความคล้ายคลึง 3 ลำดับแรก กลุ่ม N ด้านประเพณีวัฒนธรรมที่โดดเด่น	76
ตารางที่ 16 แสดงจำนวนเอกสารในกลุ่มเอกสาร E-N	77
ตารางที่ 17 แสดงผลการตรวจคำตอบหรือประเมินความถูกต้องด้วยเมตริกซ์วัดประสิทธิภาพ	78

สารบัญรูปภาพ

	หน้า
รูปที่ 1 แนวโน้มการท่องเที่ยวโดยชุมชน (Community-based tourism: CBT)	3
รูปที่ 2 ขอบเขตการวิจัยเชิงพื้นที่การท่องเที่ยวโดยชุมชนเขตอีสานใต้.....	4
รูปที่ 3 สถานการณ์ปัญหา/คำถามวิจัย/ผลลัพธ์และความคาดหวัง	8
รูปที่ 4 กรอบแนวคิดการวิจัย	12
รูปที่ 5 กรอบแนวคิดการวิจัยวัตถุประสงค์ที่ 1	13
รูปที่ 6 กรอบแนวคิดการวิจัยในวัตถุประสงค์ข้อที่ 2	17
รูปที่ 7 กรอบแนวคิดการวิจัยตามวัตถุประสงค์ที่ 3.....	22
รูปที่ 8 รูปแบบการวิเคราะห์องค์ประกอบ	35
รูปที่ 9 ข้อจำกัดและข้อดกเบื้องต้นของการใช้สถิติวิเคราะห์องค์ประกอบ.....	35
รูปที่ 10 การวิเคราะห์องค์ประกอบหลัก [62].....	36
รูปที่ 11 ผลการวิเคราะห์การหาความเชื่อมั่น Cronbach's alpha.....	40
รูปที่ 12 ผลลัพธ์ค่าสถิติ KMO และ Bartlett's test.....	41
รูปที่ 13 ผลลัพธ์ค่าความแปรปรวนของชุดข้อมูลทั้งหมด	41
รูปที่ 14 ผลลัพธ์ค่า Factor loading ทั้ง 4 กลุ่ม	42
รูปที่ 15 องค์ประกอบรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวอีสานใต้	43
รูปที่ 16 กระบวนการมาตรฐานที่ใช้สำหรับการทำเหมืองข้อมูล (CRISP-DM) [41].....	48
รูปที่ 17 กรอบการดำเนินงานการวิจัยตามวัตถุประสงค์ที่ 2	51
รูปที่ 18 ขั้นตอนการเตรียมข้อมูล (Data preparation) 4 กลุ่ม (Cluster)	52
รูปที่ 19 ตัวอย่างผลลัพธ์การสร้างกฎความสัมพันธ์ด้วยโปรแกรม Weka	53
รูปที่ 20 เมตริกซ์วัดประสิทธิภาพ (Confusion Matrix)	64
รูปที่ 21 กรอบแนวคิดการวิจัย (Conceptual framework)	68

รูปที่ 22 การเตรียมข้อมูลสำหรับชุดคำสอบถาม (Query collections).....	70
รูปที่ 23 ผลการทดลองการเตรียมข้อมูล (Data preprocessing).....	72
รูปที่ 24 ตัวอย่างกระบวนการสกัดคุณลักษณะเอกสาร.....	73
รูปที่ 25 เมตริกชี้วัดประสิทธิภาพ (Confusion Matrix) ชุดข้อมูลทั้ง 4 Class.....	77



บทที่ 1

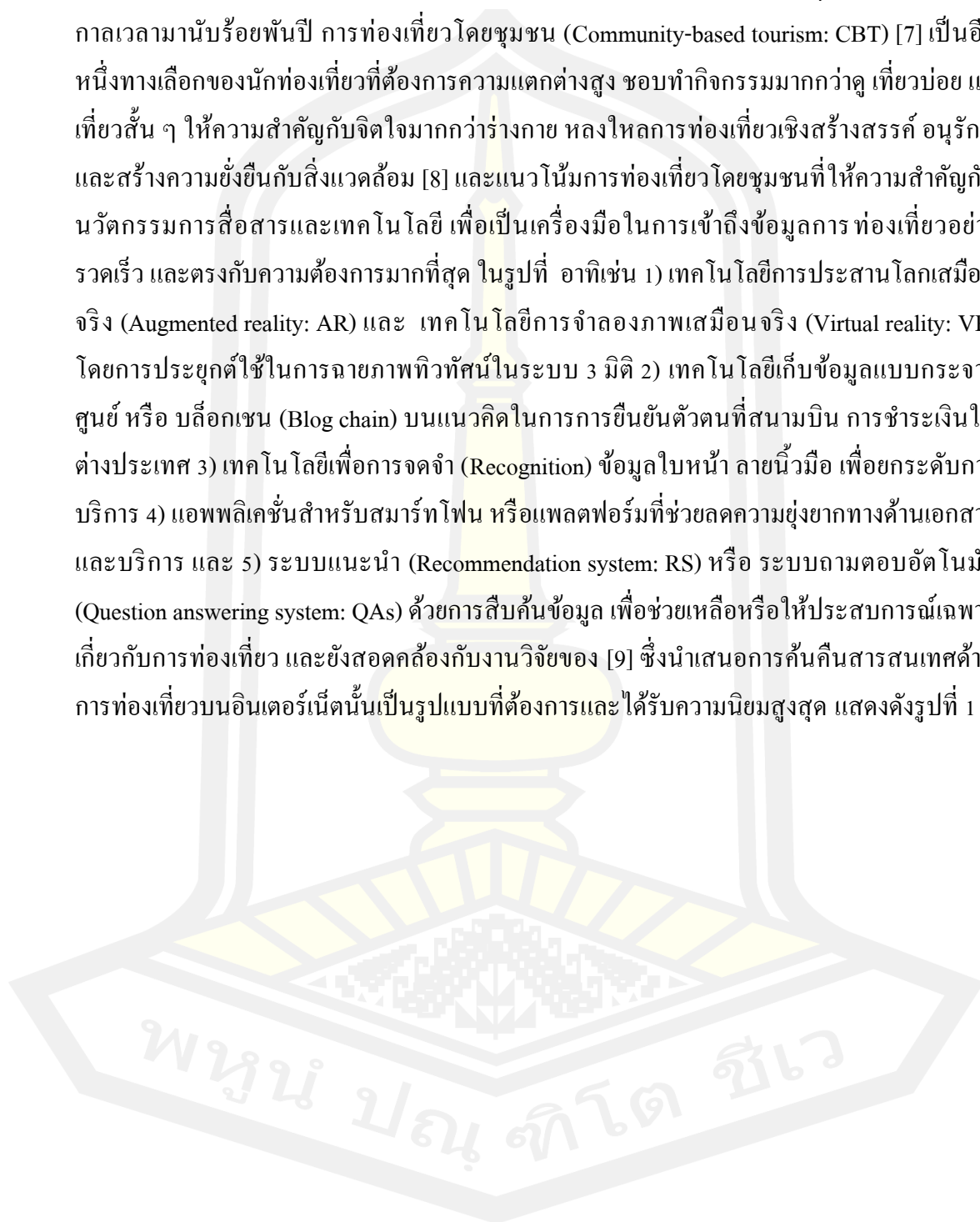
บทนำ

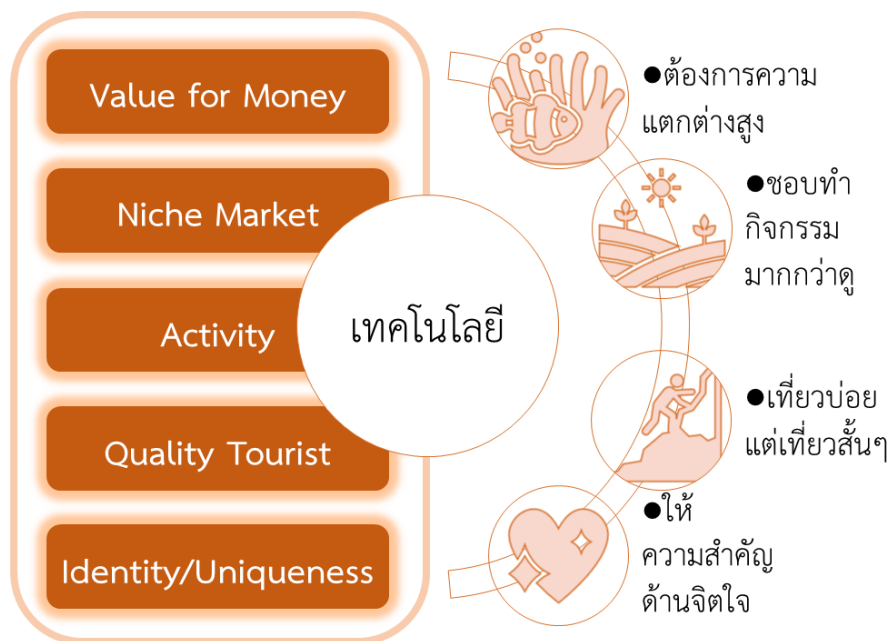
1.1 ที่มาและความสำคัญของปัญหา

เป้าหมายขององค์การสหประชาชาติ (United nations: UN) และแผนแม่บทของโลก AGENDA 21 [1] [2] [3] ด้วยการพัฒนาอย่างยั่งยืน โดยคำนึงถึงความสมดุลระหว่างการบริโภค ประชากรและความสามารถต่อการค้าจุนสิ่งมีชีวิต (Earth's life supporting capacity) รวมทั้งการพัฒนาการท่องเที่ยวทั้งหมดต้องมุ่งสู่ความยั่งยืน [4] อุตสาหกรรมท่องเที่ยวนับได้ว่าเป็นอุตสาหกรรมสำคัญที่สร้างงานให้กับประชาชนคนไทย ตลอดจนสร้างรายได้ที่เป็นเงินตราต่างประเทศมาเป็นระยะเวลายาวนาน เห็นได้จากแผนพัฒนาเศรษฐกิจและสังคมแห่งชาติ ฉบับที่ 13 (พ.ศ. 2566-2570) [5] “การท่องเที่ยวเน้นคุณค่าและความยั่งยืน” อุตสาหกรรมการท่องเที่ยวและบริการเป็นหนึ่งในยุทธศาสตร์หลักที่ภาครัฐให้ความสำคัญ “กระจายรายได้สู่การท่องเที่ยวเมืองรอง” เพื่อกระตุ้นให้เกิดการท่องเที่ยวและสร้างเศรษฐกิจฐานรากอย่างต่อเนื่อง จากสถิตินักท่องเที่ยวต่างชาติทั่วโลก ในปี 2561 เพิ่มขึ้น 6% คิดเป็นเงิน 1.4 พันล้านดอลลาร์สหรัฐ โดยทวีปเอเชียเพิ่มขึ้นมากกว่า 6% คิดเป็นเงิน 343 พันล้านดอลลาร์สหรัฐ แสดงให้เห็นการเพิ่มขึ้นของจำนวนนักท่องเที่ยวเมื่อเปรียบเทียบกับในแต่ละภูมิภาคทั่วโลก และย้อนกลับไปในปี 2560 ประเทศไทยเป็นประเทศที่สร้างรายได้จากนักท่องเที่ยวมากที่สุดในภูมิภาค 57.5 พันล้านดอลลาร์สหรัฐ ซึ่งสูงกว่าการคาดการณ์ขององค์การการท่องเที่ยวโลก (World tourism organization: UNWTO)

องค์การการท่องเที่ยวโลกได้กำหนดรูปแบบของการท่องเที่ยวใน 3 รูปแบบ ประกอบด้วย 1) การท่องเที่ยวเชิงประวัติศาสตร์ (Historical tourism) 2) การท่องเที่ยวเชิงวัฒนธรรม (Cultural and traditional tourism) และ 3) การท่องเที่ยวชนบท (Rural tourism) สอดคล้องกับยุทธศาสตร์และแนวทางพัฒนาการท่องเที่ยวเน้นคุณค่าอย่างยั่งยืน ภายใต้แนวคิดดังกล่าวกระทรวงการท่องเที่ยวและกีฬา (ททท.) ร่วมกับองค์การบริหารการพัฒนาพื้นที่พิเศษเพื่อการท่องเที่ยวอย่างยั่งยืน (อพท.) [1] ได้วางกลยุทธ์เพื่อการพัฒนาแหล่งท่องเที่ยวทางวัฒนธรรมของชุมชนบนพื้นฐานต้นทุนทางทรัพยากรที่โดดเด่น ซึ่ง “อีสานใต้” เป็นหนึ่งในดินแดนที่เต็มไปด้วยเสน่ห์ของอารยธรรมเก่าแก่ที่ได้รับความนิยมจากนักท่องเที่ยวทั่วทุกสารทิศ [6] เป็นเส้นทางท่องเที่ยวอารยธรรมขอมโบราณที่โดดเด่นเป็นอัตลักษณ์ชัดเจน เห็นได้จากร่องรอยแห่งอารยธรรมที่รุ่งโรจน์ คือกลุ่มปราสาทหินที่มีความงามเป็นที่ปรากฏหลายหนแห่ง อาทิเช่น ปราสาทหินพนมาย ในจังหวัดนครราชสีมา ปราสาทพนมรุ้ง ปราสาทเมืองต่ำ ในจังหวัดบุรีรัมย์ กลุ่มปราสาทตาเมือน ปราสาทตาควาย ในจังหวัด

สุรินทร์ และ ปราสาทสระกำแพงน้อย-ใหญ่ ปราสาทปรางกู ในจังหวัดศรีสะเกษ ซึ่งยังคงความสวยงามแห่งมนต์ลึกลับรอการต้อนรับนักท่องเที่ยวให้มาสัมผัสความงามอันทรงคุณค่าที่ยืนหยัดผ่านกาลเวลามานับร้อยพันปี การท่องเที่ยวโดยชุมชน (Community-based tourism: CBT) [7] เป็นอีกหนึ่งทางเลือกของนักท่องเที่ยวที่ต้องการความแตกต่างสูง ชอบทำกิจกรรมมากกว่าดู เที่ยวบ่อย แต่เที่ยวสั้น ๆ ให้ความสำคัญกับจิตใจมากกว่าร่างกาย หลงใหลการท่องเที่ยวเชิงสร้างสรรค์ อนุรักษ์ และสร้างความยั่งยืนกับสิ่งแวดล้อม [8] และแนวโน้มการท่องเที่ยวโดยชุมชนที่ให้ความสำคัญกับนวัตกรรมการสื่อสารและเทคโนโลยี เพื่อเป็นเครื่องมือในการเข้าถึงข้อมูลการท่องเที่ยวอย่างรวดเร็ว และตรงกับความต้องการมากที่สุด ในรูปที่ อาทิเช่น 1) เทคโนโลยีการประสานโลกเสมือนจริง (Augmented reality: AR) และ เทคโนโลยีการจำลองภาพเสมือนจริง (Virtual reality: VR) โดยการประยุกต์ใช้ในการฉายภาพทิวทัศน์ในระบบ 3 มิติ 2) เทคโนโลยีเก็บข้อมูลแบบกระจายศูนย์ หรือ บล็อกเชน (Blog chain) บนแนวคิดในการยืนยันตัวตนที่สนามบิน การชำระเงินในต่างประเทศ 3) เทคโนโลยีเพื่อการจดจำ (Recognition) ข้อมูลใบหน้า ลายนิ้วมือ เพื่อยกระดับการบริการ 4) แอปพลิเคชันสำหรับสมาร์ตโฟน หรือแพลตฟอร์มที่ช่วยลดความยุ่งยากทางด้านเอกสารและบริการ และ 5) ระบบแนะนำ (Recommendation system: RS) หรือ ระบบถามตอบอัตโนมัติ (Question answering system: QAs) ด้วยการสืบค้นข้อมูล เพื่อช่วยเหลือหรือให้ประสบการณ์เฉพาะเกี่ยวกับการท่องเที่ยว และยังคงสอดคล้องกับงานวิจัยของ [9] ซึ่งนำเสนอการค้นคืนสารสนเทศด้านการท่องเที่ยวบนอินเทอร์เน็ตนั้นเป็นรูปแบบที่ต้องการและได้รับความนิยมสูงสุด แสดงดังรูปที่ 1





รูปที่ 1 แนวโน้มการท่องเที่ยวโดยชุมชน (Community-based tourism: CBT)

แม้ประเทศไทยมีนโยบายลดทอนปัญหาความความเหลื่อมล้ำในสังคม [10] ปัญหาความยากจนยังมีให้เห็นอย่างชัดเจน จากตัวเลขการเติมโตทางเศรษฐกิจส่วนใหญ่เป็นผลประกอบการจากภาคอุตสาหกรรมและธุรกิจขนาดใหญ่ การพัฒนาอย่างก้าวกระโดดของเทคโนโลยี เป็นหนึ่งในปัจจัยสำคัญที่ก่อให้เกิดปัญหาความเหลื่อมล้ำของระบบเศรษฐกิจและการกระจายรายได้ ซึ่งคนไทยในชนบทยังมีคุณภาพชีวิตค่อนข้างต่ำ ส่งผลให้ขาดโอกาส สิทธิ การเข้าถึงทรัพยากร เทคโนโลยี รวมทั้งการรับรู้ข่าวสารอย่างรู้เท่าทันเทคโนโลยี และการเข้าถึงข้อมูลที่รวดเร็ว ทันสมัยและตรงกับความต้องการ สอดคล้องกับงานวิจัยของ [11] กล่าวถึงพฤติกรรมของนักท่องเที่ยว Gen X และ Gen Y ซึ่งส่วนใหญ่ใช้ชีวิตและเติบโตมาในช่วงยุคสมัยที่มีความก้าวหน้าทางเทคโนโลยี ให้ความสำคัญกับการติดต่อสื่อสารและความถูกต้องของข้อมูล นิยมใช้เทคโนโลยีและสื่อโซเชียลมีเดียในการเข้าถึงข้อมูลด้านการท่องเที่ยวและสิ่งดึงดูดใจ ตอบสนองความต้องการ ถูกต้องและรวดเร็ว ปัจจัยเหล่านี้ทำให้เกิดการกระตุ้นการตัดสินใจซื้อสินค้าและบริการเพื่อการท่องเที่ยวเพิ่มมากขึ้น



รูปที่ 2 ขอบเขตการวิจัยเชิงพื้นที่การท่องเที่ยวโดยชุมชนเขตอีสานใต้

ขณะที่ปัญหาสำคัญของการท่องเที่ยวโดยชุมชนในเขตอีสานใต้คือยังไม่เป็นที่รู้จักของกลุ่มนักท่องเที่ยวและขาดการประชาสัมพันธ์ให้ตรงกลุ่มเป้าหมาย แม้มีนโยบายการส่งเสริมจากภาครัฐเพื่อกระตุ้นให้เกิดการท่องเที่ยวอย่างต่อเนื่อง [12] ส่วนใหญ่เน้นการพัฒนาและส่งเสริมแหล่งท่องเที่ยวหลักที่ได้รับความนิยม (First class destination) เช่น นครราชสีมา ขณะที่แหล่งท่องเที่ยวรอง (Second class destination) มีศักยภาพภายใต้แนวคิดที่โดดเด่น อาทิเช่น บุรีรัมย์เมืองกีฬา สุรินทร์เมืองเกษตรอินทรีย์ ศรีสะเกษเมืองขอมโบราณ ชุมชนเหล่านี้ยังขาดการเข้าถึงข้อมูลและการสนับสนุนแหล่งทุนจากทุกภาคส่วน ยังขาดการส่งเสริมในเชิงพาณิชย์อย่างจริงจัง แสดงดังรูปที่ 2

การบริหารจัดการเป็นเครื่องมือที่สำคัญในการแก้ไขปัญหาเชิงพื้นที่และใช้เป็นแนวทางพัฒนาการท่องเที่ยวแบบบูรณาการอย่างยั่งยืน [13] [14] โดยการจัดการท่องเที่ยวและกระบวนการมีส่วนร่วมในการจัดการทรัพยากรอันเป็นต้นทุนทางธรรมชาติและวัฒนธรรมให้เกิดประโยชน์สูงสุดแก่ชุมชน คำนึงถึงความยั่งยืนของสิ่งแวดล้อม สังคมและวัฒนธรรม เน้นให้ชุมชนเข้ามามีบทบาทเป็นเจ้าของทรัพยากรทางการท่องเที่ยวและรับผลประโยชน์จากการจัดการทรัพยากรนั้น [7] ส่งผลต่อการกระจายตัวของเศรษฐกิจฐานราก ความเหลื่อมล้ำ การรับรู้ การเข้าถึงข้อมูลข่าวสาร การเตรียมข้อมูลเพื่อให้สารสนเทศที่รวดเร็วและตรงกับความต้องการกับนักท่องเที่ยวและผู้ให้บริการนำเที่ยว ซึ่งสอดคล้องกับยุทธศาสตร์หลักที่ภาครัฐให้ความสำคัญ ด้วยการท่องเที่ยวเน้นคุณค่าและความยั่งยืน

โดยเน้นกระจายรายได้สู่การท่องเที่ยวเมืองรอง” เพื่อกระตุ้นให้เกิดการท่องเที่ยวและสร้างเศรษฐกิจฐานรากอย่างยั่งยืน

ปัญหาสำคัญของการจัดการเทคโนโลยีสารสนเทศที่ถูกพัฒนาขึ้นในปัจจุบัน [15] คือไม่มีระบบค้นคืนที่สามารถตอบโจทย์หรือสนองต่อพฤติกรรมแสวงหาความรู้ของนักท่องเที่ยวในเชิงพื้นที่เขตอีสานใต้ได้ มากไปกว่านั้นการสืบค้นข้อมูลผ่านเครือข่ายอินเทอร์เน็ตและสื่อโซเชียลมีเดีย ด้วยโปรแกรมค้นข้อมูล (Search engine) ได้มาซึ่งผลลัพธ์คือข้อมูลจำนวนมากที่ขาดการเชื่อมโยงอย่างเป็นระบบ และข้อจำกัดที่นักท่องเที่ยวพบในการเข้าถึงหรือสืบค้นข้อมูลใน 4 ประเด็น ดังนี้ 1) การดึงและแสดงผลข้อมูลที่ไม่เกี่ยวข้องตามที่ต้องการ 2) ผลลัพธ์ที่ได้จากการสืบค้นข้อมูลมีจำนวนมาก ขาดต่อการค้นหาและนำข้อมูลไปใช้ประโยชน์ 3) นักท่องเที่ยวไม่ทราบเหตุผลของผลลัพธ์ที่โปรแกรมสืบค้นนำมาแสดงทำให้ขาดการตรวจสอบ และ 4) ผลลัพธ์จากการค้นหาขาดความแม่นยำและไม่ตรงกับความต้องการ ดังนั้นการพัฒนาเครื่องมือที่จะช่วยการสืบค้นข้อมูลจำนวนมากที่มีอยู่ จึงจำเป็นต้องมีคุณสมบัติพิเศษในการจัดกลุ่มองค์ความรู้ด้านการท่องเที่ยวเฉพาะพื้นที่เขตอีสานใต้ที่เหมาะสมอย่างมีประสิทธิภาพ โดยการนำโมเดลค้นคืนสารสนเทศ (Information retrieval) ที่แสดงรูปแบบในเชิงตรรกะเพื่อจำลองตัวเอกสาร คลังข้อมูล เอกสาร ข้อสารสนเทศที่ผู้ใช้ต้องการหรือคำสอบถาม รวมถึงการเทียบเคียงจับคู่เพื่อหาผลลัพธ์ในการค้นคืนสารสนเทศด้านการท่องเที่ยว ให้ได้ตรงตามความต้องการของนักท่องเที่ยว ด้วยแบบจำลองปริภูมิเวกเตอร์

การค้นคืนสารสนเทศ [16] [17] เป็นระบบจัดการและประมวลผลสารสนเทศในหลายรูปแบบ จัดเก็บ บันทึก ดึงข้อมูลหรือเอกสารที่ต้องการ และนำเสนอหรือจัดแสดงในหลายรูปแบบ มีหลักการค้นคืนโดยใช้ลักษณะเด่นของเนื้อหาและทำการดึงเอาข้อมูลหรือเอกสารที่ต้องการ ด้วยวิธีการแยกความเกี่ยวข้อง (Relevant) และไม่เกี่ยวข้อง (Non-Relevant) สามารถแบ่งการพัฒนา ระบบค้นคืนได้ 4 ขั้นตอน ดังนี้ 1) การเตรียมข้อมูล (Data pre-processing) 2) กระบวนการสกัดคุณลักษณะของคำ (Feature extraction) 3) การเปรียบเทียบ (Similarity matching) ระหว่างเอกสารคำถามและเอกสารคำตอบ และ 4) การประเมินประสิทธิภาพ (Model evaluation) ซึ่งแสดงความสามารถในการเข้าถึงสารสนเทศให้ตรงกับความต้องการมากที่สุด ปัจจุบัน ถูกนำมาประยุกต์ใช้เพื่อการจัดการความรู้ (Knowledge management) อย่างแพร่หลาย โดยเฉพาะระบบค้น

คืนสารสนเทศ (Information retrieval) ด้านการท่องเที่ยว (Tourism domain) ที่ง่ายต่อการใช้งาน และมีประสิทธิภาพ

การท่องเที่ยวเป็นอุตสาหกรรมภาคบริการมีบทบาทสำคัญในระบบเศรษฐกิจของโลก และประเทศไทย สร้างมูลค่าเป็นอันดับหนึ่งของสินค้าและบริการ ในขณะที่เทคโนโลยีสารสนเทศ และแนวโน้มการท่องเที่ยวมีการขยายตัวอย่างรวดเร็ว ทรัพยากรธรรมชาติและสิ่งแวดล้อมก็เสื่อมสภาพลงอย่างรวดเร็วและต่อเนื่องเช่นกัน จึงเป็นที่มาของ AGENDA 21 คู่การท่องเที่ยวอย่างยั่งยืน สอดคล้องกับยุทธศาสตร์และแนวทางพัฒนาการท่องเที่ยวเน้นคุณค่าอย่างยั่งยืน คู่แนวโน้มการท่องเที่ยวโดยชุมชน (Community-based tourism: CBT) [8] ตอกย้ำให้ทุกภาคส่วนเกิดการตื่นตัวและตระหนักถึงการจัดการทรัพยากรและสิ่งแวดล้อมให้มีความสมดุลสามารถใช้ประโยชน์ได้อย่างต่อเนื่องและยั่งยืน แต่เทคโนโลยีที่ถูกพัฒนาขึ้นไม่สามารถตอบสนองต่อพฤติกรรม การแสวงหาความรู้ของนักท่องเที่ยวในเชิงพื้นที่ได้ และมากกว่านั้นคือ ข้อมูลจำนวนมากจากการสืบค้นผ่านเครือข่ายอินเทอร์เน็ตไม่ตรงกับความต้องการ ยากต่อการค้นหา การตรวจสอบและนำข้อมูลไปใช้ประโยชน์ ดังนั้นเครื่องมือในการค้นคืนสารสนเทศจึงจำเป็นต้องมีคุณสมบัติพิเศษในการค้นคืนแยกแยะเอกสารสนเทศ รวมทั้งการจัดเรียงความสำคัญและการจัดกลุ่มองค์ความรู้ เพื่อสนับสนุนการค้นคืนสารสนเทศสำหรับการท่องเที่ยวเฉพาะพื้นที่ ให้ตรงตามความต้องการของนักท่องเที่ยวและผู้มีส่วนได้เสีย ด้วยหลักการค้นคืนสารสนเทศ (Information retrieval: IR) ได้อย่างมีประสิทธิภาพ อาทิเช่น Hamoud และคณะ [18] ใช้แนวคิดการค้นคืนสารสนเทศในคลังข้อมูลคัมภีร์อัลกุระอ่าน (Quran corpus) ประมวลผลด้วยวิธี StringToWordVector นำไปใช้ประโยชน์เพื่อพัฒนาระบบถามตอบได้ในอนาคต สอดคล้องกับ Laddha และคณะ [19] ได้พัฒนาระบบค้นคืนข้อมูลการท่องเที่ยวโดยใช้ฐานข้อมูลคำศัพท์ (WordNet) ภาษาอังกฤษ ประเมินระบบจากคำค้น (keyword-based) พบว่าระบบที่พัฒนาขึ้นง่ายต่อการใช้งาน การค้นคืนมีความรวดเร็ว แม่นยำและมีประสิทธิภาพสูง เช่นเดียวกับ Ghosh และคณะ [20] พัฒนาระบบค้นหาแหล่งท่องเที่ยวในประเทศอินเดีย ด้วยเทคนิค Human-centric faceted approach นำเสนอผลการทดลองด้วย GraphDB ผลจากการวิจัยพบว่า >90% ระบบสามารถสืบค้นและให้ข้อมูลได้ถูกต้อง แม่นยำ ตามความต้องการของนักท่องเที่ยว

ความท้าทายของงานวิจัยในครั้งนี้ คือการสร้างนวัตกรรมและตอบโจทย์กลไกการแก้ปัญหาในพฤติกรรมแสวงหาความรู้ของนักท่องเที่ยวในเชิงพื้นที่และข้อมูลจำนวนมากจากการสืบค้นผ่านเครือข่ายอินเทอร์เน็ต ซึ่งได้ผลลัพธ์ไม่ตรงกับความต้องการ ยากต่อการค้นหา

การตรวจสอบและนำข้อมูลไปใช้ประโยชน์ มากไปกว่านั้นยังช่วยให้นักท่องเที่ยวใช้วางแผนและการตัดสินใจเลือกเดินทางท่องเที่ยวด้วยตนเองได้อย่างเป็นรูปธรรม ขณะเดียวกันชุมชนและผู้ให้บริการนำเที่ยว รวมทั้งผู้มีส่วนได้เสียตลอดห่วงโซ่ สามารถใช้วางแผนบริหารจัดการทรัพยากรให้เพียงพอกับจำนวนนักท่องเที่ยว เพื่อการสร้างประสบการณ์ท่องเที่ยวแปลกใหม่ ที่อยู่ในความสนใจจะสร้างความประทับใจให้นักท่องเที่ยวกลับมาเที่ยวซ้ำและบอกต่อ ซึ่งจะส่งผลดีและเกิดประโยชน์ต่อชุมชนอย่างยั่งยืน [2] [3] [7] โดยการออกแบบและพัฒนา “การพัฒนากรอบการทำงานการหาความสัมพันธ์คำถามคำตอบอัตโนมัติสำหรับท่องเที่ยวชุมชน” เพื่อพัฒนาศักยภาพการให้บริการและการเข้าถึงข้อมูลที่มีประสิทธิภาพ รวมทั้งยกระดับมูลค่าและคุณค่าให้การท่องเที่ยวโดยชุมชน เตรียมข้อมูลสารสนเทศให้ตรงกับความต้องการของนักท่องเที่ยว สื่อสารประชาสัมพันธ์และเพิ่มประสิทธิภาพในการเข้าถึงกลุ่มเป้าหมาย ด้วยระบบการสืบค้นข้อมูลให้ตรงกับความต้องการ รวดเร็ว ถูกต้องและแม่นยำ สร้างความมั่นใจและโอกาสต่อการตัดสินใจเดินทางของนักท่องเที่ยว [7] [11] [9] [21]

1.2 โจทย์วิจัย (Research question)

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง รวมทั้งการลงพื้นที่อย่างต่อเนื่องในชุมชนท่องเที่ยวอีสานใต้ 4 จังหวัด ประกอบด้วย นครราชสีมา บุรีรัมย์ สุรินทร์ และ ศรีสะเกษ ทำให้ผู้วิจัยได้ทราบถึงปัญหาและอุปสรรคที่ส่งผลกระทบต่อการท่องเที่ยวโดยชุมชนในเขตอีสานใต้ ในงานวิจัยฉบับนี้ผู้วิจัยจึงได้ชี้ให้เห็นถึงประเด็นปัญหาที่มักจะเกิดขึ้นกับการท่องเที่ยวโดยชุมชนในเขตอีสานใต้ แสดงดังรูปที่ 3 รายละเอียดต่อไปนี้

- 1) องค์ประกอบของกลุ่มรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ควรเป็นอย่างไร?
- 2) รูปแบบรายการคำถามที่มีความสำคัญต่อการท่องเที่ยวโดยชุมชนอีสานใต้ควรเป็นอย่างไร?
- 3) คำตอบสำหรับตอบคำถามอัตโนมัติสำหรับการท่องเที่ยวโดยชุมชนอีสานใต้ควรเป็นอย่างไร? จากรายการรูปแบบคำถามอัตโนมัติสามารถพยากรณ์คำตอบล่วงหน้าได้หรือไม่?



สถานการณ์ปัญหา

- แหล่งท่องเที่ยวไม่เป็นที่รู้จัก
- ขาดการส่งเสริมอย่างจริงจัง
- การบริหารจัดการที่ล้มเหลว
- ข้อมูลไม่ถูกค้นคืน

คำถามวิจัย

- 1) องค์ประกอบของกลุ่มรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ควรเป็นอย่างไร?
- 2) รูปแบบรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ควรเป็นอย่างไร?
- 3) คำตอบสำหรับตอบคำถามอัตโนมัติสำหรับการท่องเที่ยวโดยชุมชนอีสานใต้ควรเป็นอย่างไร?

ผลลัพธ์ที่คาดหวัง

- พัฒนารอบการทำงานการหาคำตอบสำหรับตอบคำถามอัตโนมัติการท่องเที่ยวโดยชุมชนอีสานใต้ที่เหมาะสมอย่างมีประสิทธิภาพและตรงกับความต้องการ

รูปที่ 3 สถานการณ์ปัญหา/คำถามวิจัย/ผลลัพธ์และความคาดหวัง

1.3 จุดมุ่งหมายของการวิจัย

เพื่อพัฒนารอบการทำงานการหาความสัมพันธ์คำถามคำตอบอัตโนมัติสำหรับท่องเที่ยวโดยชุมชนอีสานใต้

1.4 ขอบเขตการวิจัย

งานวิจัยในครั้งนี้ได้ดำเนินการศึกษารวบรวมข้อมูล แนวคิดและทฤษฎีที่เกี่ยวข้อง นำสู่การวิเคราะห์ข้อมูลด้านการท่องเที่ยว (Tourism domain) โดยชุมชนอีสานใต้ แบ่งการดำเนินงานเป็น 3 ส่วน ประกอบด้วย 1) เพื่อวิเคราะห์องค์ประกอบของกลุ่มรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ 2) เพื่อค้นหารูปแบบคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ และ 3) เพื่อพัฒนารอบการทำงานการหาคำตอบสำหรับตอบคำถามอัตโนมัติการท่องเที่ยวโดยชุมชนอีสานใต้

1.5 วิธีการดำเนินวิจัย

ส่วนที่ 1 ศึกษารวบรวมข้อมูล แนวคิดและทฤษฎีที่เกี่ยวข้อง นำสู่การวิเคราะห์ข้อมูลด้านการท่องเที่ยว (Tourism domain) โดยชุมชนอีสานใต้ 4 จังหวัด ประกอบด้วย นครราชสีมา บุรีรัมย์ สุรินทร์ และ ศรีสะเกษ โดยวิธีการการคำนวณค่าทางสถิติด้วยการวิเคราะห์องค์ประกอบเชิงสำรวจ

(Exploratory factor analysis: EFA) เพื่อให้ได้ความสัมพันธ์ของโครงสร้างของตัวแปร โดยเลือกวิธีการสกัดองค์ประกอบ (Principal component analysis: PCA) ที่สามารถอธิบายความผันแปรของชุดข้อมูลได้มากที่สุด [22]

ส่วนที่ 2 ประยุกต์ใช้กฎความสัมพันธ์ (Association rule mining: ARM) ในการค้นหาพฤติกรรมและรูปแบบการแสวงหาความรู้ของนักท่องเที่ยว (Tourist information-seeking behaviours) ด้วย Apriori algorithm [23]

ส่วนที่ 3 การพัฒนากรอบการทำงานการหาความสัมพันธ์คำถามคำตอบอัตโนมัติสำหรับท่องเที่ยวชุมชน บนหลักการค้นคืนสารสนเทศ (Information retrieval: IR) ด้วยแบบจำลองปริภูมิเวกเตอร์ ของนักท่องเที่ยวให้สามารถเข้าถึงสารสนเทศให้ตรงกับความต้องการมากที่สุด และประเมินผลการค้นคืนสารสนเทศด้วยค่า Precision Recall และ F1 score

1.6 ผลที่คาดว่าจะได้รับ

- 1) ได้ทราบถึงปัญหาและอุปสรรคที่ส่งผลกระทบต่อการท่องเที่ยวโดยชุมชนอีสานใต้
- 2) ได้กลุ่มรายการคำถามที่มีอิทธิพลต่อการตัดสินใจเดินทางทางนักท่องเที่ยว
- 3) ได้กฎความสัมพันธ์ของรายการคำถามที่นักท่องเที่ยวถามบ่อย
- 4) ผู้ให้บริการนำเที่ยวสามารถเตรียมข้อมูลเพื่อให้สารสนเทศได้อย่างถูกต้องและรวดเร็ว
- 5) นักท่องเที่ยวสามารถเข้าถึงสารสนเทศได้ตรงกับความต้องการมากที่สุด
- 6) เกิดการใช้ประโยชน์จากทรัพยากรท้องถิ่นให้คุ้มค่าในเชิงเศรษฐกิจ สังคม วัฒนธรรม และสิ่งแวดล้อมอย่างยั่งยืน

1.7 นิยามศัพท์

1) การท่องเที่ยวเชิงวัฒนธรรม (Cultural-based tourism) หมายถึง การเดินทางไปยังแหล่งท่องเที่ยวทางประวัติศาสตร์อันทรงคุณค่าในชุมชนอีสานใต้ เพื่อศึกษาหาความรู้ เยี่ยมชมวัฒนธรรม จารีต ประเพณี วิถีชีวิตและภูมิปัญญาที่มีความเป็นเอกลักษณ์ที่โดดเด่นของชุมชนและผู้คนในท้องถิ่น สะท้อนให้เห็นความเจริญรุ่งเรืองจากอดีตจนถึงปัจจุบัน บนพื้นฐานการมีส่วนร่วมของคนในชุมชนในการใช้ประโยชน์จากทรัพยากรของท้องถิ่นให้คุ้มค่าในเชิงเศรษฐกิจ สังคม วัฒนธรรม และสิ่งแวดล้อมอย่างยั่งยืน [24]

2) การท่องเที่ยวโดยชุมชน (Community-based tourism) หมายถึง การท่องเที่ยวโดยชุมชนอีสานใต้ที่คำนึงถึงความยั่งยืนของสิ่งแวดล้อม สังคม วัฒนธรรม [24] เกิดการเรียนรู้และ

สร้างประสบการณ์ใหม่ให้แก่ผู้มาเยือนจากการใช้ประโยชน์ทรัพยากรของท้องถิ่นให้คุ้มค่า เพื่อให้คนในชุมชนมีคุณภาพชีวิตที่ดีขึ้น [1] โดยชุมชนเป็นเจ้าของและมีส่วนร่วมในการบริหารจัดการ [8]

3) อีสานใต้ หมายถึง 4 จังหวัด ได้แก่ นครราชสีมา บุรีรัมย์ สุรินทร์ และ ศรีสะเกษ

4) ESAN model หมายถึง องค์ประกอบของรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนเขตอีสานใต้ แบ่งได้ 4 กลุ่ม (Cluster) ประกอบด้วย (1) การให้บริการที่เป็นเลิศ (Excellent service) (2) สิ่งอำนวยความสะดวกที่มีมาตรฐาน (Standard facilities) (3) ข้อมูลสารสนเทศและประชาสัมพันธ์ (Accuracy of information) และ (4) อัตลักษณ์ ประเพณีและวัฒนธรรม (Noble culture)

5) Question answering system (Q&As) หมายถึง การพัฒนากรอบการทำงานการหาความสัมพันธ์คำถามคำตอบอัตโนมัติสำหรับท่องเที่ยวชุมชน

6) การค้นคืนสารสนเทศ (Information retrieval: IR) หมายถึง ความสามารถของนักท่องเที่ยวในการเข้าถึงสารสนเทศให้ตรงกับความต้องการมากที่สุด

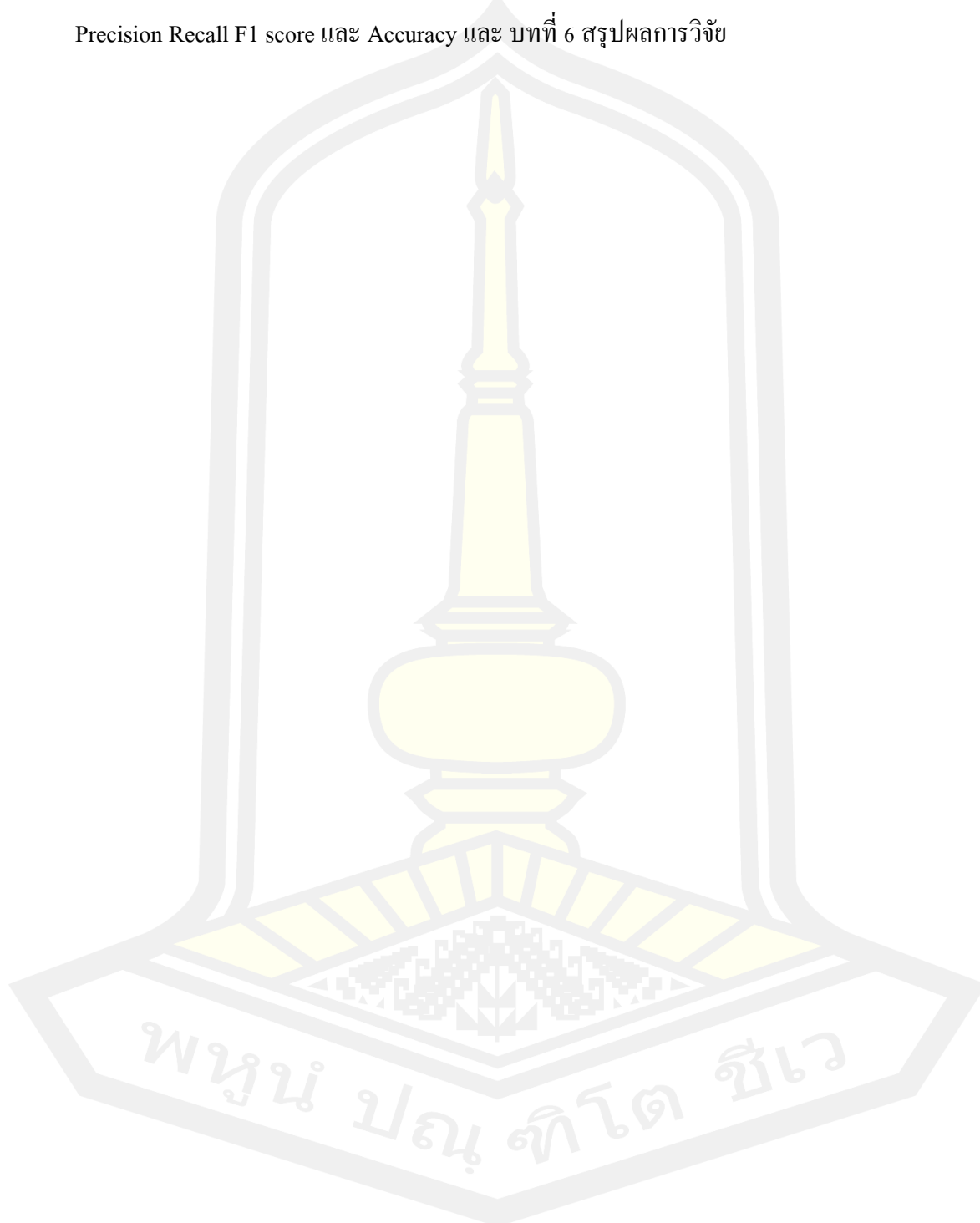
7) โมเดลการค้นคืนสารสนเทศ (Information retrieval model) หมายถึง ต้นแบบที่เหมาะสมสำหรับการประเมินประสิทธิภาพการค้นคืน

8) การประเมินผล (Evaluate) หมายถึง การประเมินประสิทธิภาพและประสิทธิผลของระบบถามตอบอัตโนมัติ จากการค้นคืนสารสนเทศด้วยค่า Precision Recall F1 score และ Accuracy

1.8 รายละเอียดของวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้ ผู้วิจัยได้นำเสนอเนื้อหาใน 6 บทที่สำคัญ ประกอบด้วย บทที่ 1 บทนำ กล่าวถึงความเป็นมาและความสำคัญของปัญหา โจทย์วิจัย จุดมุ่งหมายของการวิจัย ขอบเขตของงานวิจัย วิธีการดำเนินงานวิจัย ผลที่คาดว่าจะได้รับ นิยามศัพท์ และ รายละเอียดของเค้าร่างวิทยานิพนธ์ บทที่ 2 วิธีการดำเนินงานวิจัย แนวคิดพื้นฐานและขั้นตอนการดำเนินงาน โดยแบ่งตาม 3 วัตถุประสงค์ของการวิจัยและบทความที่ได้รับการเผยแพร่ในระดับนานาชาติ จำนวน 3 บทความ บทที่ 3 การวิเคราะห์องค์ประกอบของกลุ่มรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ ด้วยวิธีคำนวณค่าทางสถิติวิเคราะห์การองค์ประกอบเชิงสำรวจ (Exploratory factor analysis: EFA) โดยวิธีการสกัดองค์ประกอบ (Principal component analysis: PCA) บทที่ 4 การค้นหารูปแบบคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ โดยการวิเคราะห์กฎความสัมพันธ์ (Association rule mining: ARM) ด้วยอัลกอริทึม Apriori บทที่ 5 การพัฒนากรอบการทำงานการหาคำตอบสำหรับตอบคำถามอัตโนมัติการท่องเที่ยวโดยชุมชนอีสานใต้ ด้วยแนวคิดการค้นคืนข้อมูลจากเอกสารและข้อความ (Textual information retrieval) บนพื้นฐานแบบจำลอง

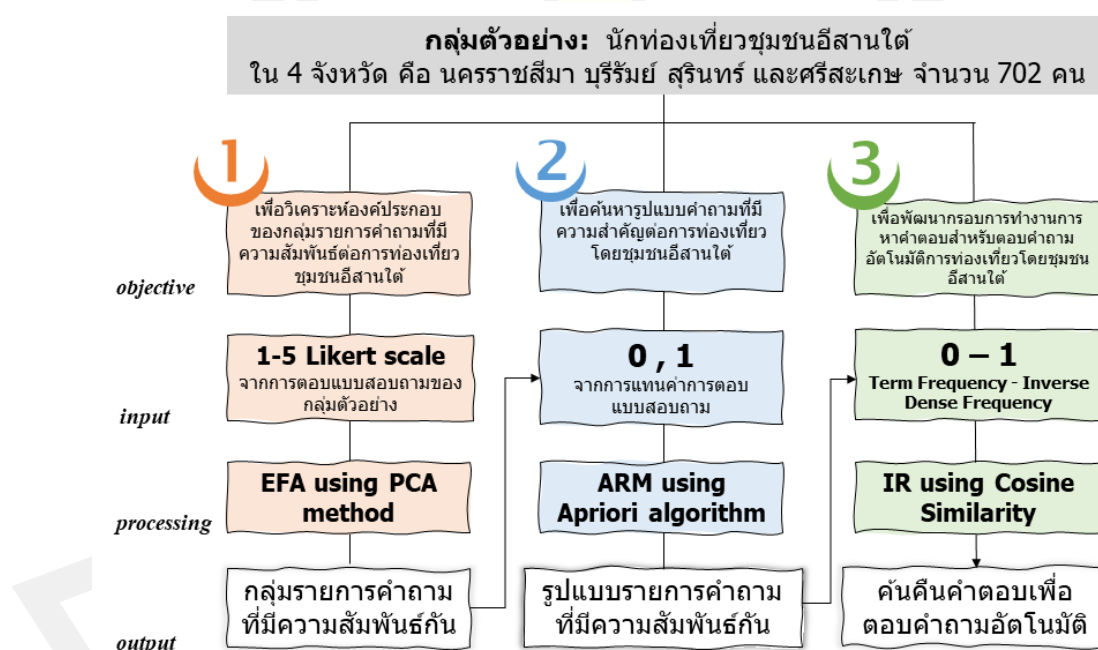
Vector space model โดยคำนวณหาค่าความคล้ายคลึงจากการวัดระยะห่างด้วยเทคนิค Cosine similarity อธิบายคำตอบและประสิทธิภาพโมเดลที่สร้างขึ้นด้วยแบบจำลองทางคณิตศาสตร์ด้วยค่า Precision Recall F1 score และ Accuracy และ บทที่ 6 สรุปผลการวิจัย



บทที่ 2

วิธีการดำเนินงานวิจัย

งานวิจัยในครั้งนี้ได้ดำเนินการศึกษารวบรวมข้อมูล แนวคิดและทฤษฎีที่เกี่ยวข้อง นำสู่การวิเคราะห์ข้อมูลด้านการท่องเที่ยว (Tourism domain) โดยชุมชนอีสานใต้ โดยผู้วิจัยได้ชี้ให้เห็นถึงประเด็นปัญหาที่มักจะเกิดขึ้นกับการท่องเที่ยวโดยชุมชนในเขตอีสานใต้ใน 3 ประเด็นคำถามวิจัย คือ 1) องค์ประกอบของกลุ่มรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ควรเป็นอย่างไร? 2) รูปแบบรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ควรเป็นอย่างไร? และ 3) คำตอบสำหรับตอบคำถามอัตโนมัติสำหรับการท่องเที่ยวโดยชุมชนอีสานใต้ควรเป็นอย่างไร? บนพื้นฐานการดำเนินงานและกรอบแนวคิดการวิจัยในภาพรวม แสดงดังรูปที่ 4 และจัดลำดับการนำเสนอตามวัตถุประสงค์ดังรายละเอียดต่อไปนี้



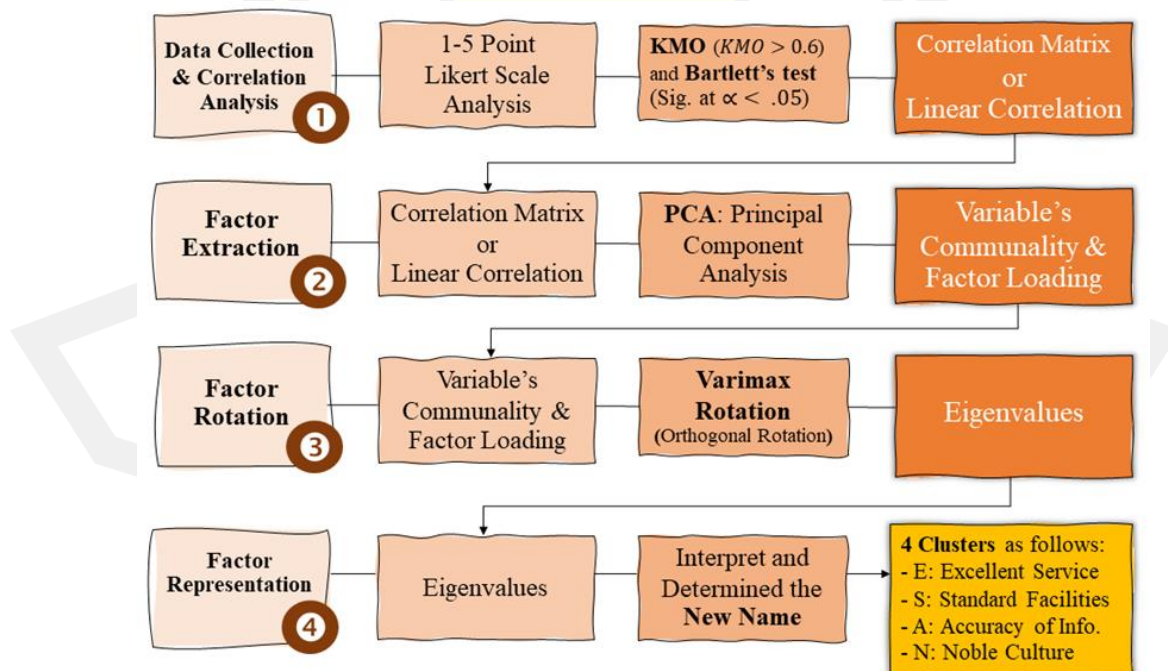
รูปที่ 4 กรอบแนวคิดการวิจัย

2.1 การวิเคราะห์องค์ประกอบของกลุ่มรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอิสานใต้

ดำเนินการศึกษารวบรวมข้อมูล แนวคิดและทฤษฎีที่เกี่ยวข้อง นำสู่การวิเคราะห์ข้อมูลด้านการท่องเที่ยว (Tourism domain) โดยชุมชนอิสานใต้ 4 จังหวัด ประกอบด้วย นครราชสีมา บุรีรัมย์ สุรินทร์ และ ศรีสะเกษ โดยวิธีการการคำนวณค่าทางสถิติด้วยการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory factor analysis: EFA) เพื่อให้ได้ความสัมพันธ์ของโครงสร้างของตัวแปร โดยเลือกวิธีการสกัดองค์ประกอบ (Principal component analysis: PCA) ที่สามารถอธิบายความผันแปรของชุดข้อมูลได้มากที่สุด [22]

2.1.1 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ดำเนินการศึกษาแนวคิดและทฤษฎีที่เกี่ยวข้อง ประกอบไปด้วย 1) 5A ปัจจัยสำคัญของการท่องเที่ยว (5A's Component of tourism) [25] 2) พฤติกรรมนักท่องเที่ยว (Tourist behaviours) [26] 3) แรงจูงใจการท่องเที่ยว (Travel motivation) [27] [28] 4) 4P ส่วนผสมทางการตลาด (4P's Marketing mix) และ 5) หุ่นยนต์สนทนาและระบบถามตอบอัตโนมัติ (Chatbot or QAs) นำสู่การสร้างแบบสอบถามและตรวจสอบความเหมาะสมของแบบสอบถาม โดยผู้เชี่ยวชาญ ก่อนการนำไปใช้งานจริง กระบวนการดังรูปที่ 5



รูปที่ 5 กรอบแนวคิดการวิจัยวัตถุประสงค์ที่ 1

2.1.2 การกำหนดขอบเขตปัญหาของงานวิจัย

จากการลงพื้นที่อย่างต่อเนื่องในชุมชนท่องเที่ยวอีสานใต้ 4 ในจังหวัด ได้แก่ นครราชสีมา บุรีรัมย์ สุรินทร์ และ ศรีสะเกษ ทำให้ผู้วิจัยได้ทราบถึงปัญหาและอุปสรรคที่ส่งผลกระทบต่อการท่องเที่ยวโดยชุมชนในเขตอีสานใต้ ในงานวิจัยฉบับนี้ผู้วิจัยจึงได้ชี้ให้เห็นถึงประเด็นปัญหาที่มักจะเกิดขึ้นกับการท่องเที่ยวโดยชุมชนในเขตอีสานใต้คือ “องค์ประกอบของกลุ่มรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ควรเป็นอย่างไร?”

2.1.3 การเก็บรวบรวมข้อมูล

จากการลงพื้นที่ทั้ง 4 จังหวัดอย่างต่อเนื่อง งานวิจัยในครั้งนี้ใช้วิธีการเก็บรวบรวมข้อมูลด้วยแบบสอบถามออนไลน์ จากกลุ่มตัวอย่างนักท่องเที่ยวของเครือข่ายการท่องเที่ยวโดยชุมชนอีสานใต้ใน 4 จังหวัด ได้แก่ นครราชสีมา บุรีรัมย์ สุรินทร์ และศรีสะเกษ ดำเนินการเก็บรวบรวมข้อมูล รวมทั้งสิ้น 702 คน/ตัวอย่าง โดยให้นักท่องเที่ยวตอบแบบสอบถามจาก Google form ผ่านอุปกรณ์อัจฉริยะ (Smart device) จำนวน 3 ตอน รวมทั้งสิ้น 67 ข้อ/รายการคำถาม ซึ่งคำถามเป็นแบบเลือกตอบและแบบมาตราส่วนประมาณค่า 5 ระดับ เมื่อได้ผลลัพธ์จากการตอบแบบสอบถามของนักท่องเที่ยว ผู้วิจัยเลือกใช้ข้อมูลจากผลลัพธ์ในตอนที่ 3 รายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ โดยเลือกให้ความสำคัญกับน้ำหนักหรือคะแนนในระดับมาก (4) และ มากที่สุด (5) เพียง 2 ระดับเท่านั้น จากนั้นนำข้อมูลไปวิเคราะห์ด้วยโปรแกรมสำเร็จรูปทางสถิติ เพื่อวิเคราะห์องค์ประกอบเชิงสำรวจในขั้นตอนต่อไป

1) ประชากรและกลุ่มตัวอย่าง เป็นนักท่องเที่ยวชุมชนอีสานใต้ใน 4 จังหวัด ได้แก่ นครราชสีมา บุรีรัมย์ สุรินทร์ และศรีสะเกษ ระหว่างเดือนมกราคม 2562 ถึง ตุลาคม 2563 จำนวน 702 คน กำหนดกลุ่มตัวอย่างด้วยแนวคิดของ Yamane [29] ที่ความเชื่อมั่น 95% และความคลาดเคลื่อนที่ 5% ดังสมการ 2.1

สมการ

$$n = \frac{N}{1 + Ne^2} \quad (2.1)$$

เมื่อ

n แทน ขนาดของกลุ่มตัวอย่าง

N แทน ขนาดของประชากร

e แทน ความคลาดเคลื่อนของการสุ่มตัวอย่าง

2) การสร้างเครื่องมือที่ใช้ในงานวิจัย เริ่มจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง และ นำสู่สร้างข้อคำถามตามกรอบแนวคิดการวิจัย โดยแบ่งแบบสอบถามออกเป็น 3 ตอน ประกอบด้วย

ตอนที่ 1 ข้อมูลทั่วไปของนักท่องเที่ยว เป็นข้อมูลพื้นฐานของผู้ตอบแบบสอบถามเพื่อใช้ สำหรับการอ้างอิงแหล่งข้อมูล จำนวน 6 ข้อคำถาม ได้แก่ เพศ อายุ สถานภาพ การศึกษาสูงสุด รายได้ และอาชีพ ซึ่งคำถามเป็นแบบเลือกตอบซึ่งเลือกได้คำตอบเดียว

ตอนที่ 2 พฤติกรรมของนักท่องเที่ยว เป็นข้อมูลพฤติกรรม ลักษณะและรูปแบบการเดินทาง ของนักท่องเที่ยว จำนวน 11 ข้อคำถาม ซึ่งคำถามเป็นแบบเลือกตอบแบ่งเป็น 1) เลือก ได้คำตอบ เดียว และ 2) เลือกได้หลายคำตอบ โดยผู้วิจัยได้ระบุไว้ในข้อคำถามอย่างชัดเจน ตัวอย่างข้อคำถาม เช่น วัตถุประสงค์ การเดินทาง ที่พัก ระยะเวลา และรูปแบบการท่องเที่ยว ฯลฯ

ตอนที่ 3 รายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้เป็นข้อมูล การแสดงความคิดเห็นจากนักท่องเที่ยวชุมชนอีสานใต้ จำนวน 50 รายการคำถาม ลักษณะของ แบบสอบถามแบบมาตราส่วนประมาณค่า 5 ระดับ [30] ลิเคิร์ต (Likert scale) ซึ่งผู้วิจัยกำหนด น้ำหนักหรือคะแนนในการทำแบบสอบถามดังนี้

รายการคำถามมีความสัมพันธ์ในระดับมากที่สุด	ให้น้ำหนักหรือคะแนนเป็น 5
รายการคำถามมีความสัมพันธ์ในระดับมาก	ให้น้ำหนักหรือคะแนนเป็น 4
รายการคำถามมีความสัมพันธ์ในระดับปานกลาง	ให้น้ำหนักหรือคะแนนเป็น 3
รายการคำถามมีความสัมพันธ์ในระดับน้อย	ให้น้ำหนักหรือคะแนนเป็น 2
รายการคำถามมีความสัมพันธ์ในระดับน้อยที่สุด	ให้น้ำหนักหรือคะแนนเป็น 1

3) ตรวจสอบความตรงของเนื้อหา (Content validity) การวิจัยครั้งนี้ดำเนินการตรวจสอบ ความตรงของเนื้อหาด้วยค่าดัชนีความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์ (Item-objective congruence index: IOC) จากผู้เชี่ยวชาญ จำนวน 5 คน ที่มีประสบการณ์ด้านการท่องเที่ยวมากกว่า 10 ปีขึ้นไป โดยพิจารณาให้คะแนนตามเกณฑ์ แล้วนำผลคะแนนมาวิเคราะห์หาดัชนีความ สอดคล้อง [31] โดยพิจารณาเลือกข้อคำถามที่มีค่า IOC ตั้งแต่ 0.50-1.00 ถือว่ามีความเที่ยงตรงตาม เนื้อหา ทั้งนี้หากมีข้อเสนอแนะจากผู้เชี่ยวชาญจะต้องนำมาพิจารณาแก้ไขให้สมบูรณ์ก่อนนำไป วัดผล หรือหากค่าความเชื่อมั่นต่อไป โดยดัชนีความสอดคล้อง แสดงดังสมการ 2.2

ระดับคะแนนเท่ากับ 1	คือ มั่นใจในความสอดคล้อง
ระดับคะแนนเท่ากับ 0	คือ ไม่มั่นใจในความสอดคล้อง
ระดับคะแนนเท่ากับ -1	คือ มั่นใจในความไม่ความสอดคล้อง

สมการ

$$IOC = \frac{\sum R}{n} \quad (2.2)$$

เมื่อ $\sum R$ คือ ผลคูณของคะแนนกับจำนวนผู้เชี่ยวชาญ
 n คือ จำนวนผู้เชี่ยวชาญ

4) การหาความเชื่อมั่นแอลฟาของครอนบาค (Cronbach's alpha reliability coefficient) [32] ของแบบสอบถามด้วยการทดสอบค่าสัมประสิทธิ์แอลฟาของครอนบาค เป็นหนึ่งในวิธีการหาความเชื่อมั่นแบบความคงที่ภายใน (Internal consistency) เพื่อหาค่าความเชื่อมั่นหรือความเที่ยงตรงของแบบสอบถาม [33] ยอมรับค่าสัมประสิทธิ์แอลฟา .70 ขึ้นไป เพื่อตรวจสอบความเข้าใจของผู้ตอบแบบสอบถาม (Pilot survey) โดยผู้วิจัยดำเนินการทดลองเก็บข้อมูลจากกลุ่มตัวอย่างนักท่องเที่ยวทดลองเที่ยว (FAM trip) ในจังหวัดสุรินทร์ จำนวน 36 คน มีค่าความเชื่อมั่นแอลฟาของครอนบาค .983 แสดงดังสมการที่ 2.3

สมการ

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum v_i}{v_t} \right) \quad (2.3)$$

เมื่อ α คือ ค่าความเชื่อมั่น
 k คือ จำนวนข้อคำถาม
 $\sum v_i$ คือ ผลรวมความแปรปรวนของคะแนนแต่ละข้อ
 v_t คือ ความแปรปรวนของคะแนนรวม

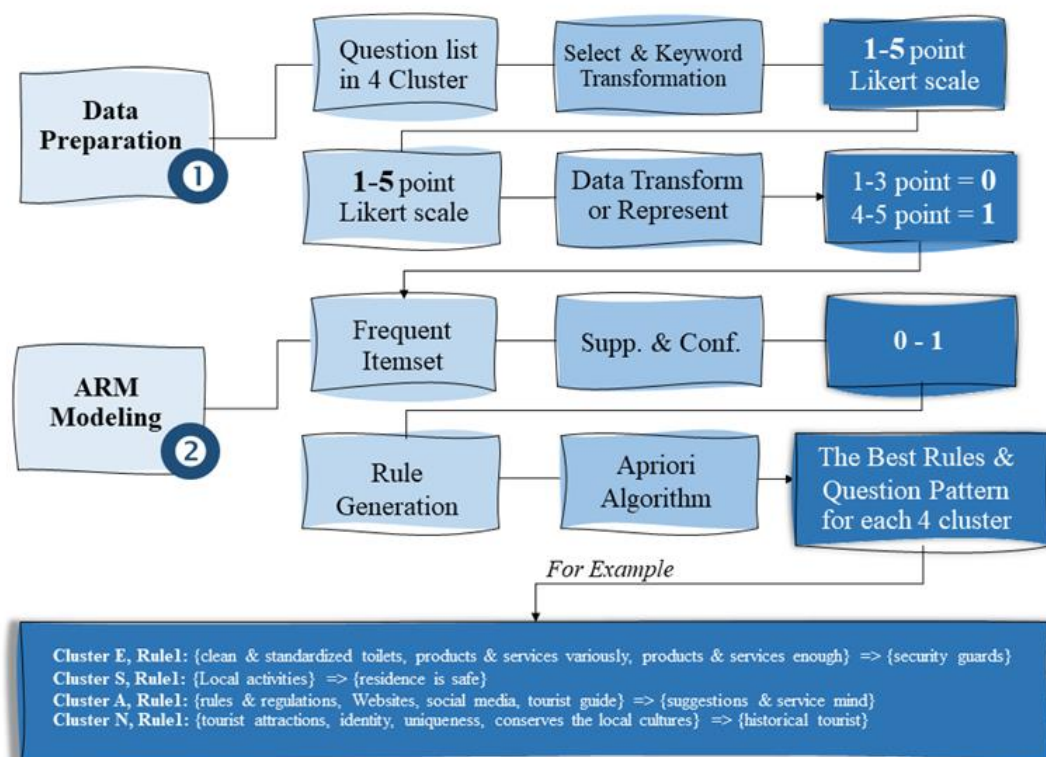
2.1.4 วิธีการวิเคราะห์ข้อมูล

การวิจัยในครั้งนี้ ผู้วิจัยเลือกใช้ข้อมูลจากผลลัพธ์จากแบบสอบถามในตอนต้นที่ 3 วิเคราะห์ข้อมูลด้วยโปรแกรมสำเร็จรูปทางสถิติ จากรายการคำถามที่มีอิทธิพลต่อการตัดสินใจเดินทางนักท่องเที่ยว คำนวณค่าทางสถิติด้วยการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory factor analysis: EFA) เพื่อให้ได้ความสัมพันธ์ของโครงสร้างของตัวแปร โดยเลือกวิธีการสกัดองค์ประกอบ (Principal component analysis: PCA) ที่สามารถอธิบายความผันแปรของชุดข้อมูลได้มากที่สุด เลือกวิธีการหมุนแกนองค์ประกอบแบบตั้งฉาก (Varimax orthogonal rotation) สำหรับเกณฑ์ที่ใช้ในการตัดสินใจเลือกตัวแปรที่สำคัญจากองค์ประกอบที่เกิดขึ้นจากการหมุนแกนต้องมีค่าสัมประสิทธิ์ (Factor score coefficients) ตั้งแต่ 0.5 ขึ้นไป รายละเอียดนำเสนอในบทที่ 3

2.2 การค้นหารูปแบบคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้

2.2.1 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ดำเนินการศึกษาแนวคิดและทฤษฎีที่เกี่ยวข้อง ประกอบไปด้วย 1) แนวคิดและทฤษฎีด้านการท่องเที่ยว (Tourism concepts) [24] [34] [35] [25] [15] [17] 2) การแสวงหาความรู้ (Information-seeking) [38] 3) กฎความสัมพันธ์ (Association rule mining: ARM) และ 4) ระบบถามตอบอัตโนมัติ (Question answering system: QAs) [21] [39] ขั้นตอนการวิจัยในรูปที่ 6



รูปที่ 6 กรอบแนวคิดการวิจัยในวัตถุประสงค์ข้อที่ 2

2.2.2 การกำหนดขอบเขตปัญหาของงานวิจัย

ในขั้นตอนนี้ผู้วิจัยได้กำหนดขอบเขตปัญหาของงานวิจัยโดยศึกษารูปแบบพฤติกรรมกรรมการสอบถามและลำดับความสัมพันธ์การสอบถาม สำหรับการท่องเที่ยวในพื้นที่อีสานใต้ ซึ่งผู้วิจัยได้ชี้ให้เห็นถึงประเด็นและทำความเข้าใจปัญหา (Business understanding) ที่มักจะเกิดขึ้นกับการท่องเที่ยวโดยชุมชนในเขตอีสานใต้ของวัตถุประสงค์นี้ คือ "เพื่อค้นหารูปแบบคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้" โดยนำวิธีการเทคนิคเหมืองข้อมูล (Data mining: DM) ตามกรอบแนวคิดการวิเคราะห์เหมืองข้อมูล CRISP-DM และการประยุกต์ใช้กฎ

ความสัมพันธ์ (ARM) ในการวิเคราะห์รายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ ด้วยอัลกอริทึม Apriori ในโปรแกรม Weka เวอร์ชัน 3.6 สำหรับค้นหารูปแบบการสอบถามหรือการแสวงหาข้อมูลของนักท่องเที่ยวที่ปรากฏร่วมกันบ่อย ๆ (Frequent item set)

2.2.3 การเก็บรวบรวมข้อมูล (Data collection)

1) การเก็บรวบรวมข้อมูล ผู้วิจัยใช้ข้อมูลเดิมจากวัตถุประสงค์ข้อ 1 รวมทั้งสิ้น 702 คน/ตัวอย่าง 50 รายการคำถาม ด้วยแบบสอบถามออนไลน์จากกลุ่มตัวอย่างนักท่องเที่ยวของเครือข่ายการท่องเที่ยวโดยชุมชนอีสานใต้ใน 4 จังหวัดเช่นเดิม การวิเคราะห์ข้อมูลในครั้งนี้ ทำความเข้าใจข้อมูล (Data understanding) ตรวจสอบคุณภาพและการคัดเลือกข้อมูลที่จะนำมาวิเคราะห์เป็นสิ่งสำคัญ ผู้วิจัยเลือกใช้ข้อมูลจากผลลัพธ์ในตอนต้นที่ 3 รายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ โดยเลือกให้ความสำคัญกับน้ำหนักหรือคะแนนในระดับมาก (4) และ มากที่สุด (5) เพียง 2 ระดับเท่านั้น

2) การเตรียมข้อมูล (Data preparation) ผู้วิจัยเลือกใช้ข้อมูลผลการตอบแบบสอบถามในตอนต้นที่ 3 จำนวน 50 รายการคำถาม จากกลุ่มตัวอย่างผู้ตอบแบบสอบถามรวม 702 คน โดยแบบสอบถามที่ใช้แบ่งเป็น 3 ตอน ประกอบด้วย ตอนที่ 1 ข้อมูลทั่วไป ตอนที่ 2 พฤติกรรมการท่องเที่ยว และตอนที่ 3 รายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวอีสานใต้ ซึ่งในขั้นตอนนี้ ผู้วิจัยเลือกนำข้อมูลเฉพาะตอนที่ 3 มาวิเคราะห์ โดยพิจารณาให้ความสำคัญกับน้ำหนักหรือคะแนนจากการตอบแบบสอบถามเพียง 2 ระดับ คือ ระดับมาก (4) และ มากที่สุด (5) ซึ่งการใช้อัลกอริทึม Apriori ในโปรแกรม Weka มีข้อจำกัดคือ ข้อมูลที่นำมาใช้จะอยู่ในรูปแบบ Nominal หรือ Ordinal เท่านั้น ในขั้นตอนการเตรียมข้อมูลดำเนินการ โดยการเลือกนำข้อมูลการตอบแบบสอบถามเฉพาะตอนที่ 3 มาวิเคราะห์ โดยแทนค่าข้อมูลน้ำหนักคะแนน 1-3 ด้วยเลข 0 และแทนค่าข้อมูลน้ำหนักคะแนน 4-5 ด้วยเลข 1 จัดเก็บและบันทึกข้อมูลลงในโปรแกรมไมโครซอฟท์เอ็กเซล ในรูปแบบไฟล์ .CSV แสดงรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวอีสานใต้ ดังตารางที่ 1

ระดับน้อยที่สุด (1) ระดับน้อย (2) และ ระดับปานกลาง (3)	แทนค่าด้วย	0
ระดับมาก (4) และ ระดับมากที่สุด (5) แทน	แทนค่าด้วย	1

ตารางที่ 1 รายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวอิสานใต้ 50 รายการคำถาม

Question no.	Question list
Q01	ชุมชนมีแหล่งท่องเที่ยวที่หลากหลายมั๊ย?
Q02	ชุมชนมีแหล่งท่องเที่ยวที่มีอัตลักษณ์ (Identity) เฉพาะท้องถิ่นมั๊ย?
Q03	ชุมชนมีแหล่งท่องเที่ยวที่เป็นเอกลักษณ์ (Uniqueness) โดดเด่นมั๊ย?
Q04	ชุมชนมีแหล่งท่องเที่ยวทางประวัติศาสตร์ที่สะท้อนวัฒนธรรม วิถีชีวิต และประเพณีมั๊ย?
Q05	ชุมชนมีการอนุรักษ์วัฒนธรรมท้องถิ่นอย่างเข้มแข็งมั๊ย?
Q06	การเดินทางไปยังแหล่งท่องเที่ยวเข้าถึงได้สะดวกมั๊ย?
Q07	เส้นทางคมนาคมในการเข้าถึงแหล่งท่องเที่ยวอยู่ในสภาพดีมั๊ย?
Q08	มีป้ายบอกเส้นทางในการเข้าถึงแหล่งท่องเที่ยวที่ชัดเจนและมีจำนวนที่เหมาะสมมั๊ย?
Q09	ระยะเวลาในการเดินทางไปยังแหล่งท่องเที่ยวมีความเหมาะสมมั๊ย?
Q10	มียานพาหนะที่หลากหลายรูปแบบในการเดินทางเข้าถึงแหล่งท่องเที่ยวมั๊ย?
Q11	มียานพาหนะสำหรับการเดินทางท่องเที่ยวในพื้นที่ เช่น จักรยาน สามล้อ อีแต่น เกวียน ฯลฯ มั๊ย?
Q12	มีที่พักเพียงพอต่อความต้องการมั๊ย?
Q13	มีสิ่งอำนวยความสะดวกภายในห้องพัก มั๊ย?
Q14	ที่พักมีความสะอาดมั๊ย?
Q15	เข้าถึงที่พักได้อย่างสะดวกมั๊ย?
Q16	ที่พักมีความปลอดภัยมั๊ย?
Q17	ที่พักมีเจ้าหน้าที่/เจ้าของบ้านคอยให้ความช่วยเหลือตลอดเวลามั๊ย?
Q18	ที่พักมีราคาที่เหมาะสมมั๊ย?
Q19	ความหลากหลายของประเภทที่พักมั๊ย?
Q20	กิจกรรมการท่องเที่ยวมีความน่าสนใจมั๊ย?
Q21	กิจกรรมการท่องเที่ยวมีความหลากหลายมั๊ย?
Q22	มีกิจกรรมการท่องเที่ยวที่เปิดโอกาสให้นักท่องเที่ยวได้มีส่วนร่วมมั๊ย?
Q23	เวลาในการดำเนินกิจกรรมการท่องเที่ยวมีความเหมาะสมมั๊ย?
Q24	กิจกรรมการท่องเที่ยวสามารถสะท้อนเอกลักษณ์วัฒนธรรม วิถีชีวิต และภูมิปัญญาท้องถิ่นมั๊ย?
Q25	กิจกรรมการท่องเที่ยวสร้างความรู้ความเข้าใจในเอกลักษณ์ของชุมชนมั๊ย?
Q26	ชุมชนมีพื้นที่รองรับนักท่องเที่ยวอย่างเพียงพอมั๊ย?
Q27	ประชาชนในท้องถิ่นให้การบริการและสนับสนุนด้านการท่องเที่ยวเป็นอย่างดีมั๊ย?
Q28	แหล่งท่องเที่ยวมีสิ่งอำนวยความสะดวกพื้นฐานเพียงพอมั๊ย?
Q29	แหล่งท่องเที่ยวมีสัญญาณ โทรศัพท์และสัญญาณอินเทอร์เน็ต ครอบคลุมอย่างทั่วถึงมั๊ย?
Q30	แหล่งท่องเที่ยวมีสถานบริการด้านสุขภาพมั๊ย?
Q31	แหล่งท่องเที่ยวมีสถานที่จอดรถที่เหมาะสมและปลอดภัยมั๊ย?
Q32	แหล่งท่องเที่ยวมีการจัดระเบียบจราจรได้อย่างเหมาะสมมั๊ย?
Q33	แหล่งท่องเที่ยวมีห้องน้ำที่สะอาด มีมาตรฐาน และเพียงพอต่อจำนวนนักท่องเที่ยวมั๊ย?
Q34	แหล่งท่องเที่ยวมีร้านอาหารที่สะอาด ได้มาตรฐาน และเพียงพอต่อจำนวนนักท่องเที่ยวมั๊ย?
Q35	มีอาหารที่เป็นเอกลักษณ์เฉพาะของชุมชนบริการแก่นักท่องเที่ยวมั๊ย?
Q36	แหล่งท่องเที่ยวมีร้านอาหารนำสินค้าชุมชน ของฝากและของที่ระลึกมั๊ย?
Q37	ร้านค้ากำหนดราคาสินค้าและบริการได้อย่างเหมาะสมมั๊ย?
Q38	ร้านค้ามีสินค้าและบริการที่หลากหลายมั๊ย?
Q39	ร้านค้ามีสินค้าและบริการที่มีคุณภาพได้มาตรฐานมั๊ย?
Q40	ร้านค้ามีสินค้าและบริการที่เพียงพอแก่นักท่องเที่ยวมั๊ย?
Q41	แหล่งท่องเที่ยวมีเจ้าหน้าที่ดูแลความปลอดภัยสำหรับนักท่องเที่ยวมั๊ย?
Q42	แหล่งท่องเที่ยวมีมัคคุเทศก์/ผู้นำเที่ยวท้องถิ่นที่มีความรู้มั๊ย?
Q43	แหล่งท่องเที่ยวมีการให้บริการด้านข้อมูลและมีศูนย์ข้อมูลแก่นักท่องเที่ยวมั๊ย?
Q44	แหล่งท่องเที่ยวแจ้งระเบียบและข้อปฏิบัติให้นักท่องเที่ยวทราบอย่างชัดเจนมั๊ย?
Q45	เจ้าหน้าที่ได้คำแนะนำ การบริการที่ดีและมีคุณภาพมั๊ย?
Q46	แหล่งท่องเที่ยวมีระบบสารสนเทศและเว็บไซต์ให้ข้อมูลแก่นักท่องเที่ยวมั๊ย?
Q47	แหล่งท่องเที่ยวมีช่องทางสื่อสารกับนักท่องเที่ยว เช่น เฟสบุ๊ก ไลน์ มั๊ย?
Q48	มีการประชาสัมพันธ์กิจกรรมการท่องเที่ยวผ่าน Social media อย่างต่อเนื่องมั๊ย?
Q49	มีคู่มือหรือเอกสารเผยแพร่การท่องเที่ยว และมีรายละเอียดต่าง ๆ ครบถ้วนมั๊ย?
Q50	มัคคุเทศก์/ผู้นำเที่ยวท้องถิ่นให้บริการข้อมูลที่ดีและมีคุณภาพมั๊ย?

เข้าสู่กระบวนการแปลงรูปข้อมูล (Data transformation) ด้วยการจัดเก็บและบันทึกข้อมูลลงในโปรแกรมไมโครซอฟท์เอ็กเซล ให้อยู่ในรูปแบบไฟล์ .CSV เพื่อให้สามารถนำข้อมูลมาวิเคราะห์หาความสัมพันธ์ (Association rule) ด้วยอัลกอริทึม Apriori ในโปรแกรม Weka แสดงการแทนค่าข้อมูลดังตารางที่ 2-3

ตารางที่ 2 การแทนค่าข้อมูลน้ำหนักระแนด้วยเลข 0 และ 1 (การเตรียมข้อมูล)

ผู้ตอบแบบสอบถาม (Respondent)	รายการคำถาม (Question list)																
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	...	Q44	Q45	Q46	Q47	Q48	Q49	Q50
1	4	5	4	4	5	4	5	5	5	...	4	5	4	4	5	5	5
2	5	4	4	5	4	5	5	4	4	...	5	4	4	5	4	5	5
3	5	5	5	5	4	5	5	4	5	...	4	4	5	4	4	4	5
4	4	3	5	2	4	4	5	4	4	...	4	4	4	4	5	4	4
5	3	4	4	5	2	2	3	1	1	...	3	1	1	3	1	1	3
6	5	4	5	5	4	5	4	5	5	...	5	4	5	4	5	5	4
7	2	3	5	4	5	3	5	1	2	...	3	1	2	4	2	2	1
8	4	5	5	4	5	5	4	5	5	...	4	5	4	5	5	4	5
9	4	3	5	3	5	4	5	4	5	...	5	4	5	4	5	5	5
10	3	5	4	5	3	2	3	3	4	...	2	3	2	1	2	3	2
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
702	4	5	5	4	4	5	5	4	5	...	4	5	5	4	4	4	4

ตารางที่ 3 ข้อมูลการตอบแบบสอบถามที่ถูกแทนค่าของทั้ง 4 กลุ่ม (การเตรียมข้อมูล)

ผู้ตอบแบบสอบถาม (Respondent)	รายการคำถาม (Question list)																
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	...	Q44	Q45	Q46	Q47	Q48	Q49	Q50
1	1	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1
4	1	0	1	0	1	1	1	1	1	...	1	1	1	1	1	1	1
5	0	1	1	1	0	0	0	0	1	...	0	0	0	0	0	0	0
6	1	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1
7	0	0	1	1	1	0	1	0	0	...	0	0	0	1	0	0	0
8	1	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1
9	1	0	1	0	1	1	1	1	1	...	1	1	1	1	1	1	1
10	0	1	1	1	0	0	0	0	1	...	0	0	0	0	0	0	0
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
702	1	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1

จากตารางที่ 2-3 แสดงค่าคะแนนจากการตอบแบบสอบถามของกลุ่มตัวอย่างในตอนต้นที่ 3 เป็นแบบ 5 ระดับคะแนน (Likert scale) โดยแทนค่าข้อมูลน้ำหนักระแนด้วยเลข 0 และแทนค่าข้อมูลน้ำหนักระแน 4-5 ด้วยเลข 1 บันทึกข้อมูลในโปรแกรมไมโครซอฟท์เอ็กเซล และแปลงให้อยู่ในรูปแบบไฟล์ .CSV

2.2.4 วิธีการวิเคราะห์ข้อมูลพ้องเขม

การวิจัยในครั้งนี้ได้ประยุกต์ใช้กฎความสัมพันธ์ เพื่อค้นหารูปแบบรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ โดยวิธีการกฎความสัมพันธ์ ด้วย Apriori algorithm [23] นำเข้าข้อมูลจากการตอบแบบสอบถามของนักท่องเที่ยวที่มีรูปแบบการสอบถามปรากฏร่วมกันบ่อย (Frequent item sets) จำนวน 702 เรคคอร์ด 50 รายการคำถาม แบ่งการวิเคราะห์ข้อมูลออกเป็น 4 กลุ่ม (Clusters) นำสู่การสร้างกฎความสัมพันธ์และค้นหารูปแบบรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยว โดยกำหนดค่าเริ่มต้นในโปรแกรม Weka 3.6 (minSupport = 0.1, minMetric = 0.9, numRules = 100) ดำเนินการวิเคราะห์ข้อมูลทีละกลุ่ม จำนวน 4 ครั้ง ได้กฎความสัมพันธ์รวมทั้งสิ้น 400 กฎ ในที่นี้ผู้วิจัยได้พิจารณาเลือกกฎความสัมพันธ์ที่ดีที่สุด (Best rules) เพื่อใช้เป็นข้อมูลนำเข้าในวัตถุประสงค์ข้อที่ 3 โดยพิจารณาประเด็นสำคัญใน 2 ขั้นตอนดังนี้

1) พิจารณาเลือกกฎความสัมพันธ์ของตัวชี้วัดประสิทธิภาพจากค่า confidence ที่ยอมรับได้ตั้งแต่ 98% ขึ้นไป

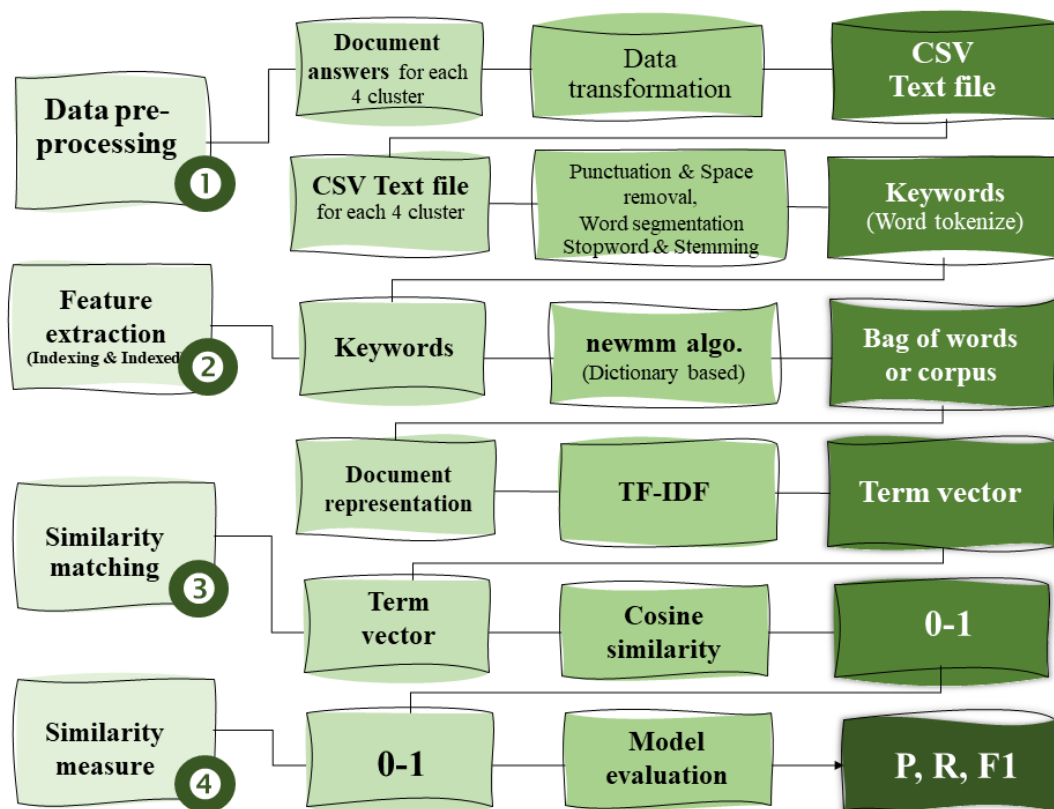
2) พิจารณาเลือกกฎที่มีลักษณะเด่น มีความเฉพาะที่เป็นไปในทิศทางเดียวกันในเชิงประจักษ์และสอดคล้องกับแนวคิด 5 ลำดับขั้นความต้องการของ Maslow โดยตัดกฎที่มีความซ้ำซ้อนทิ้ง [40]

พอสรุปผลการวิเคราะห์กฎความสัมพันธ์รายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวในพื้นที่การท่องเที่ยวโดยชุมชนเขตอีสานใต้ เพื่อนำผลลัพธ์จากขั้นตอนนี้ไปเป็น ข้อสอบถามหรือข้อมูลนำเข้าในวัตถุประสงค์ข้อที่ 3 โดยพิจารณาจากคำสำคัญ (Keywords) ของทั้ง 4 กลุ่ม โดยสรุปประกอบด้วย 1) ด้านการบริการที่เป็นเลิศ (Excellence service) 2) ด้านสิ่งอำนวยความสะดวกที่มีมาตรฐาน (Standard facilities) 3) ข้อมูลสารสนเทศที่ถูกต้องและตรงกับความต้องการ (Accuracy of information) และ 4) วัฒนธรรมที่โดดเด่น (Noble culture) เพื่อการพัฒนากรอบการทำงานการหาความสัมพันธ์คำถามคำตอบอัตโนมัติสำหรับท่องเที่ยวชุมชน ซึ่งเป็นวัตถุประสงค์สุดท้ายของงานวิจัย โดยรายละเอียดจะนำเสนอในบทที่ 4

2.3 การพัฒนากรอบการทำงานการหาคำตอบสำหรับตอบคำถามอัตโนมัติการท่องเที่ยวโดยชุมชนอีสานใต้

2.3.1 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ดำเนินการศึกษาแนวคิดและทฤษฎีที่เกี่ยวข้อง ประกอบไปด้วย 1) เหมืองข้อมูล (Data mining) [41] 2) การค้นคืนสารสนเทศ (Information retrieval) [42] และ 3) ระบบถามตอบอัตโนมัติ (Question answering system: QAs) [21] [39] โดยขั้นตอนการวิจัยนำเสนอแสดงในรูปที่ 9



รูปที่ 7 กรอบแนวคิดการวิจัยตามวัตถุประสงค์ที่ 3

2.3.2 การกำหนดขอบเขตปัญหาของงานวิจัย

จุดเริ่มต้นที่สำคัญคือสภาพปัญหาของ Search engine และ Document ที่มีอยู่ในปัจจุบัน อยากรู้ว่าการสืบค้นหรือค้นคืน ขณะเดียวกันข้อมูลที่ได้จากการสืบค้นไม่ตรงกับความต้องการของนักท่องเที่ยวนำสู่คำถามวิจัย “คำตอบสำหรับตอบคำถามอัตโนมัติสำหรับการท่องเที่ยวโดยชุมชนอีสานใต้ควรเป็นอย่างไร?” ในขั้นตอนนี้ผู้วิจัยจึงได้กำหนดขอบเขตปัญหาของงานวิจัยโดยให้ระบบสามารถจัดการและประมวลผลสารสนเทศในหลายรูปแบบ จัดเก็บ บันทึก ดึงข้อมูลหรือเอกสารที่ต้องการและนำเสนอ มีหลักการค้นคืนโดยใช้ลักษณะเด่นของเนื้อหาและทำการดึงเอาข้อมูลหรือเอกสารที่ต้องการ ด้วยวิธีการแยกความเกี่ยวข้อง (Relevant) และไม่เกี่ยวข้อง (Non-relevant) สามารถแบ่งการพัฒนาระบบค้นคืนได้ 4 ขั้นตอน ดังนี้ 1) การเตรียมข้อมูล (Data pre-processing) 2) กระบวนการสกัดคุณลักษณะของคำ (Feature extraction) 3) การเปรียบเทียบระหว่างเอกสารคำถามและเอกสารคำตอบ (Similarity matching) และ 4) การประเมินประสิทธิภาพ (Model evaluation) ทั้งนี้โมเดลที่พัฒนาขึ้นจำเป็นต้องประเมินประสิทธิภาพ (Efficiency) และประสิทธิผล (Effectiveness) ก่อนนำไปใช้งานจริง

2.3.3 การเก็บรวบรวมข้อมูล

การรวบรวมข้อมูลนำเข้าทั้งหมดเป็นข้อความภาษาไทย (Thai text documents) ที่มีเนื้อหาเกี่ยวกับการท่องเที่ยวโดยชุมชนในเขตอีสานใต้ ในขั้นตอนนี้มีข้อมูลนำเข้าจาก 2 แหล่ง คือ

1) รายการคำถามหรือข้อสอบถาม จากกฎความสัมพันธ์รูปแบบพฤติกรรมกรรมการแสวงหาความรู้ของนักท่องเที่ยว รวมทั้งสิ้น 14 กฎ จาก 4 กลุ่ม (Cluster) ซึ่งเป็นผลลัพธ์จากวัตถุประสงค์ที่ 2 โดยนำกฎทั้งหมดไปผ่านกระบวนการ Data pre-processing ผลลัพธ์ที่ได้คือ ชุดข้อมูลในรูปแบบคำสำคัญ (Keywords) ที่เกิดจากพฤติกรรมกรรมการสอบถามที่ปรากฏร่วมกันบ่อยของนักท่องเที่ยว ซึ่งจะนำมาใช้เป็นคำถาม หรือข้อสอบถามสำหรับการค้นคืนเอกสารในการหาความสัมพันธ์คำถามคำตอบอัตโนมัติสำหรับท่องเที่ยวชุมชนต่อไป

2) ข้อมูลด้านการท่องเที่ยวอีสานใต้ใน 4 จังหวัด ประกอบด้วย นครราชสีมา บุรีรัมย์ สุรินทร์ และ ศรีสะเกษ ทั้งที่เป็นข้อมูลปฐมภูมิและทุติยภูมิ อาทิเช่น Search engine ผ่านเครือข่ายอินเทอร์เน็ต ดำเนินการรวบรวมและจัดเก็บไฟล์ข้อมูลให้อยู่ในรูปของเอกสารด้วยโปรแกรมสำเร็จรูปไมโครซอฟท์เวิร์ด เพื่อใช้เป็นคำตอบจากการใช้คำสำคัญในการค้นคืนเอกสารในการหาความสัมพันธ์คำถามคำตอบอัตโนมัติสำหรับท่องเที่ยวชุมชนต่อไป

2.3.4 วิธีการวิเคราะห์ข้อมูลพอสังเขป

การวิเคราะห์ข้อมูลในส่วนนี้ มีจุดมุ่งหมายสำคัญคือการพัฒนากรอบการทำงานการหาความสัมพันธ์คำถามคำตอบอัตโนมัติสำหรับท่องเที่ยวชุมชน เพื่อเป็นต้นแบบการค้นคืนสารสนเทศประเภทข้อความหรือตัวอักษร (Textual information retrieval) โดยนักวิจัยเลือกใช้เครื่องมือที่ได้รับความนิยมกับงานหลายด้านในปัจจุบัน อาทิเช่น ปัญญาประดิษฐ์ (Artificial intelligence: AI) และวิทยาศาสตร์ข้อมูล (Data science) ซึ่งให้บริการโดย Google colabory หรือ Colab ใช้แพ็คเกจ PythaiNLP โดยสามารถเขียนและเรียกใช้งานด้วยโปรแกรมภาษาไพธอน (Python) ทำหน้าที่เป็น Jupyter notebook บน Cloud สามารถใช้งานร่วมกันผ่าน Google Drive นำมาใช้เป็นเครื่องมือสำหรับการวิเคราะห์ข้อมูลดำเนินการบนพื้นฐานของโมเดลที่หลากหลาย โดยเลือกนำจุดเด่นของแต่ละโมเดลมาประยุกต์ใช้อย่างลงตัว ประกอบด้วย 4 ขั้นตอนสำคัญ คือ

1) การเตรียมข้อมูล (Data preprocessing) ประกอบด้วยโมเดลพื้นฐานสำหรับการทำความสะอาดข้อมูล (Text cleaning) และ การประมวลผลคำ Text processing อาทิเช่น อันได้แก่ การตัดช่องว่างและสัญลักษณ์พิเศษ (Space and punctuation removal) ตามมาด้วยการตัดคำ (Word segmentation) บนพื้นฐานของการตัดคำแบบสอดคล้องมากที่สุด (Maximal matching) หรือ

newmm ด้วยหลักการเปรียบเทียบอ้างอิงจากพจนานุกรม (Dictionary-based approach: DCB) โดยสามารถสรุปกระบวนการของ Maximal matching [43] ได้ 3 ขั้นตอนพอสังเขปดังนี้ ขั้นที่ 1 ดำเนินการบนผลลัพธ์ของ Longest Matching ขั้นที่ 2 เลือกผลลัพธ์การตัดคำที่มีจำนวนค่าน้อยที่สุด และขั้นที่ 3 หากมีจำนวนค่าเท่ากันจะย้อนกลับ (Backtracking) ไปใช้เกณฑ์ขั้นที่ 1 ในการเลือกคำที่ยาวที่สุด และเป็นวิธีการที่นำมาใช้ในการทดลองครั้งนี้ เพื่อใช้แก้ไขความบกพร่องของการตัดคำแบบเลือกคำยาวที่สุด (Longest Matching) การหยุดคำ (Stop word removal) และการหารากศัพท์ (Stemming) เพื่อเพิ่มประสิทธิภาพในการค้นคืนด้วย Porter algorithm [44] [45]

2) การสร้างดัชนีคำสำคัญ (Inverted indexing) หรือสร้างตัวแทนเอกสารด้วยค่าน้ำหนักของคำ เพื่อนำความถี่มาคำนวณหาค่าน้ำหนักของคำช่วยให้การสืบค้นที่รวดเร็วขึ้น ผลลัพธ์ที่ได้ในขั้นตอนนี้ได้แก่ 1) พจนานุกรม (Dictionary) 2) การประกาศ (Posting) เก็บไว้ในหน่วยความจำหลัก และมีตัวชี้ (Pointer) ชี้ไปยัง Posting ที่เก็บไว้ในหน่วยความจำสำรอง อย่างมีมาตรฐานและจัดเก็บในรูปของฐานข้อมูล (Corpus)

3) การตัวแทนเอกสาร (Document representation) ที่ได้จากการสกัดคุณลักษณะ (Feature extraction) โดยผลลัพธ์ที่ได้รับในขั้นตอนนี้คือ โมเดลถุงคำ (Bag of words: BoW) และ 2) ค่าความถี่ของคำร่วมกับความถี่เอกสารพหุคูณ (Term frequency-inverse document frequency: TF-IDF) หรือ Term-document matrix ซึ่งให้ความสำคัญกับความถี่ของคำ (Term frequency) และ น้ำหนักของคำ (Term weight) ที่พบในเอกสาร เพื่อคำนวณหาความคล้ายคลึงกันของเอกสารทั้งคู่นบนพื้นฐาน Vector space model: VSM [18] ในขั้นตอนต่อไป

4) การเปรียบเทียบความคล้ายระหว่างคำสอบถามกับเอกสาร โดยให้ความสำคัญกับ Term-document matrix ด้วยวิธีการทางสถิติเพื่อใช้วัดค่าความคล้ายคลึงแบบโคไซน์ (Cosine similarity) หรือการวัดระยะห่าง (Distance) ของเวกเตอร์ระหว่างเอกสาร 2 เอกสาร กรณีที่ค่า Cosine มีค่าเข้าใกล้ 1 แสดงถึงความคล้ายคลึงมีมาก หากมีค่าเข้าใกล้ 0 จะมีความคล้ายคลึงน้อย ดังนั้นค่า Cosine จะมีค่าอยู่ระหว่าง 0 ถึง 1 โดยผลลัพธ์ที่ได้จะถูกจัดลำดับจากค่าและความเหมือนระหว่างคำสอบถามกับเอกสาร เพื่อให้ได้ผลลัพธ์การค้นคืนสารสนเทศให้ได้ตรงตามความต้องการของผู้ใช้

5) การประเมินประสิทธิภาพโมเดล (Model evaluation) เป็นขั้นตอนสุดท้ายเพื่อสร้างความเชื่อมั่นก่อนนำโมเดลไปใช้งานจริง โดยสามารถในการดึงเอาเอกสารที่มีความเกี่ยวข้อง (Relevant) แยกออกจากเอกสารที่ไม่เกี่ยวข้อง (Non-relevant) จากชุดข้อมูลอย่างชัดเจน โดยผลลัพธ์ที่ปรากฏคือเอกสารที่มีความสัมพันธ์กับคำค้นของผู้ใช้และถูกจัดเรียงลำดับตามความ

ต้องการ สร้างความพึงพอใจต่อผู้ใช้งานอย่างรวดเร็ว ถูกต้องและแม่นยำ ด้วยวิธีการประเมินความสามารถของแบบจำลองการประเมินประสิทธิภาพและประสิทธิผล ตามแนวคิดการค้นคืนสารสนเทศ (Information retrieval model) ใน 2 ส่วน ได้แก่ 1) การประเมินโดยผู้เชี่ยวชาญ (Expert) จำนวน 3 ท่าน ใน 3 ด้าน ประกอบด้วย ด้านเหมืองข้อมูล (Data mining) ด้านเทคโนโลยีสารสนเทศ (Information technology) และ ด้านภาษาศาสตร์ (Linguists of Thai) และ 2) การประเมินโดยระบบ (System) เพื่อตรวจสอบความถูกต้องของการดึงเอกสารที่เกี่ยวข้อง



บทที่ 3

การวิเคราะห์องค์ประกอบกลุ่มรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดย ชุมชนอีสานใต้

การดำเนินการในวัตถุประสงค์ที่แรกของการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาเอกสารและผลงานวิจัยที่เกี่ยวข้องกับสาระสำคัญ เพื่อให้เกิดความเข้าใจที่ชัดเจน ตรงประเด็น เพื่อค้นหารายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวอีสานใต้ โดยผู้วิจัยนำเสนอใน 6 ประเด็นสำคัญ ประกอบด้วย 1) ที่มาและความสำคัญของปัญหา 2) เอกสารและงานวิจัยที่เกี่ยวข้อง 3) ระเบียบวิธีวิจัย 4) ผลการทดลอง 5) สรุปผลการทดลอง และ 6) อภิปรายผล ตามลำดับดังนี้

3.1 ที่มาและความสำคัญของปัญหา

อุตสาหกรรมท่องเที่ยวเป็นหนึ่งในกำลังสำคัญต่อการขับเคลื่อนเศรษฐกิจประเทศไทย ที่สร้างรายได้มหาศาลจากนักท่องเที่ยว โดยเป็นอุตสาหกรรมภาคบริการที่มีการเติบโตทางเศรษฐกิจในสัดส่วนที่มากกว่าภาคอุตสาหกรรมอื่น โดยมีปัจจัยสนับสนุนที่สำคัญจากเศรษฐกิจทั้งในและต่างประเทศ รวมทั้งการออกมาตรการส่งเสริมการท่องเที่ยวของภาครัฐและเอกชนดำเนินการตามแผนพัฒนาการท่องเที่ยวแห่งชาติ พ.ศ. 2555-2559 ซึ่งถือเป็นทิศทางการพัฒนาการท่องเที่ยวของประเทศไทย [13] ส่งผลให้ในปี 2560 ประเทศไทยจะมีรายได้จากการท่องเที่ยวประมาณ 2.77 ล้านบาท ขยายตัวร้อยละ 10 จากปีก่อน “การท่องเที่ยวโดยชุมชน” เป็นหนึ่งในเครื่องมือสำคัญในการพัฒนาและแข่งขันของภาคการท่องเที่ยวภายใต้ห่วงโซ่มูลค่า (Value chain) เพื่อสร้างมูลค่าเพิ่มให้สินค้าและบริการทางการท่องเที่ยว ด้วยเทคโนโลยีและนวัตกรรมบริการสมัยใหม่ สามารถเข้าถึงและตอบสนองความต้องการของนักท่องเที่ยวได้มากที่สุดในปัจจุบัน ด้วยความโดดเด่นและเสน่ห์เฉพาะถิ่นของสิ่งจังหวัดของอีสานใต้ นครราชสีมา บุรีรัมย์ สุรินทร์ และศรีสะเกษ ที่มีวัฒนธรรมท้องถิ่นเป็นเอกลักษณ์ สามารถเดินทางท่องเที่ยวได้ตลอดฤดูกาล นำสู่การการแปลงทุนทางสังคมเป็นสินทรัพย์ภายใต้การจัดการการท่องเที่ยวและสร้างกลยุทธ์ [46] [47] เพื่อเพิ่มขีดความสามารถในการแข่งขันทางการท่องเที่ยวเพื่อความยั่งยืน (Sustainable tourism) [48] พัฒนาคนและเน้นการมีส่วนร่วมของชุมชนในการบริหารจัดการทรัพยากรอย่างมีมาตรฐาน [46] [49] [50]

ปัจจุบันนักท่องเที่ยวมีพฤติกรรมและแรงจูงใจต่อการตัดสินใจเพื่อเดินทางท่องเที่ยวที่แตกต่างกัน การแสวงหาการบริการสารสนเทศและแหล่งข้อมูลสารสนเทศของนักท่องเที่ยวในยุคปัจจุบันจึงปรับเปลี่ยนไปตามนวัตกรรมดิจิทัล ซึ่งสอดคล้องกับ Digital tourism platform เครื่องมือเงินนโยบายของกระทรวงการท่องเที่ยวและกีฬา [51] ด้วยการบูรณาการทางเทคโนโลยีการ

ประมวลภาษาธรรมชาติ (Natural language processing: NLP) ซึ่งเป็นแขนงหนึ่งของปัญญาประดิษฐ์ (Artificial intelligence: AI) “Chatbot” หรือหุ่นยนต์สนทนาอัตโนมัติ ถูกพัฒนาและออกแบบภายใต้แนวคิดที่ทำให้คอมพิวเตอร์สามารถสนทนาโต้ตอบกับมนุษย์ได้อย่างเป็นธรรมชาติ การออกแบบรายการคำถามที่ดีจึงเป็นองค์ประกอบสำคัญของการออกแบบระบบหุ่นยนต์สนทนาอัตโนมัติ [36] [52] [53] [54] ในการเข้าใจภาษาธรรมชาติและความต้องการที่แท้จริงในบทสนทนาของนักท่องเที่ยว เป็นพื้นฐานที่สำคัญให้ Chatbot ได้เรียนรู้สำหรับการตอบคำถามได้อย่างถูกต้อง [55] ตรงประเด็น กระชับ ง่ายต่อความเข้าใจและสื่อสารได้อย่างรวดเร็ว ส่งผลต่อการตัดสินใจเดินทางท่องเที่ยวและเป็นฐานในการออกแบบและพัฒนาหุ่นยนต์สนทนา (Chatbot) ให้สามารถสื่อสารกับมนุษย์ผ่านช่องทางออนไลน์แบบเรียลไทม์ (Real-time) [56] ให้การบริการข้อมูลข่าวสารอย่างครอบคลุมความต้องการและเข้าถึงกลุ่มนักท่องเที่ยวได้ตลอดเวลาอย่างมีประสิทธิภาพ

งานวิจัยในครั้งนี้ โดยมีวัตถุประสงค์เพื่อวิเคราะห์องค์ประกอบของกลุ่มรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยว โดยชุมชนอีสานใต้ เพื่อนำสู่การค้นหารูปแบบคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ และเพื่อพัฒนารอบการทำงานการหาคำตอบสำหรับตอบคำถามอัตโนมัติการท่องเที่ยวโดยชุมชนอีสานใต้ในลำดับต่อไป

3.2 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้เลือกนำแนวคิดทฤษฎีมาประยุกต์ใช้ในการออกแบบรายการคำถาม เพื่อให้ได้แบบสอบถามที่เกี่ยวข้องและส่งผลต่อการตัดสินใจเดินทางของนักท่องเที่ยว ประกอบด้วย 6 แนวคิดสำคัญ ได้แก่ 1) พฤติกรรมของนักท่องเที่ยว (Tourist behavior) 2) แรงจูงใจในการท่องเที่ยว (Travel motivation) 3) ส่วนผสมทางการตลาด (4P's Marketing mix) 4) องค์ประกอบของการท่องเที่ยว (5A's Component of tourism) 5) การวิเคราะห์องค์ประกอบเชิงสำรวจ และ 6) การวิเคราะห์องค์ประกอบหลัก เพื่อนำผลขององค์ประกอบร่วมของกลุ่มรายการคำถามมาเป็นฐานและเป็นแนวทางสำหรับการออกแบบหุ่นยนต์สนทนาต่อไป

3.2.1 แนวคิดและทฤษฎีด้านการท่องเที่ยว (Tourism concept)

1) การท่องเที่ยวโดยชุมชน (Community-based tourism: CBT) เป็นรูปแบบการท่องเที่ยวทางเลือกที่มีลักษณะเฉพาะตัว ได้รับความนิยมนักท่องเที่ยวอย่างแพร่หลายในปัจจุบัน ขณะที่องค์การการท่องเที่ยวโลก [24] ได้กำหนดประเภทการท่องเที่ยวในลักษณะคล้ายกัน โดยมีชื่อเรียกที่แตกต่างกันไปหากแต่มิมีนัยสำคัญเชิงความหมายที่คล้ายคลึงกัน อาทิเช่น

1.1) การท่องเที่ยวเชิงนิเวศ (Ecotourism) หมายถึง การท่องเที่ยวในแหล่งธรรมชาติที่มีเอกลักษณ์เฉพาะถิ่นและแหล่งวัฒนธรรมที่เกี่ยวข้องกับระบบนิเวศ โดยมีกระบวนการเรียนรู้ร่วมกันของผู้ที่เกี่ยวข้องภายใต้การจัดการสิ่งแวดล้อมและการท่องเที่ยวอย่างมีส่วนร่วมของท้องถิ่น เพื่อมุ่งเน้นให้เกิดจิตสำนึกต่อการรักษาระบบนิเวศอย่างยั่งยืน

1.2) การท่องเที่ยวเชิงธรณีวิทยา (Geo-tourism) หมายถึง การท่องเที่ยวในแหล่งธรรมชาติที่เป็นหินผา ลานหินทราย อุโมงค์โพรง ถ้ำน้ำลอด ถ้ำหินงอกหินย้อย เพื่อดูความงามของภูมิทัศน์ที่มีความแปลกของการเปลี่ยนแปลงของพื้นที่โลก ศักยภาพของหิน ดิน แร่ต่างๆ และฟอสซิล ได้รับความรู้ได้มีประสบการณ์ใหม่ บนพื้นฐานการท่องเที่ยวอย่างรับผิดชอบ มีจิตสำนึกต่อการรักษา สภาพแวดล้อม โดยประชาชนในท้องถิ่นมีส่วนร่วมต่อการจัดการการท่องเที่ยว

1.3) การท่องเที่ยวเชิงประวัติศาสตร์ (Historical tourism) หมายถึงการเดินทางท่องเที่ยวไปยังแหล่งท่องเที่ยวทางโบราณคดี และประวัติศาสตร์ เพื่อชื่นชมและเพลิดเพลินในสถานที่ท่องเที่ยวได้ความรู้มีความเข้าใจต่อประวัติศาสตร์และโบราณคดี ในท้องถิ่นพื้นฐานของความรับผิดชอบต่อและมีจิตสำนึกต่อการรักษามรดกทางวัฒนธรรมและคุณค่า ของสภาพแวดล้อมโดยที่ประชาชนในท้องถิ่น มีส่วนร่วมต่อการจัดการการท่องเที่ยว

1.4) การท่องเที่ยวชมวัฒนธรรมและประเพณี (Cultural and traditional tourism) หมายถึง การเดินทางท่องเที่ยว เพื่อชมงานประเพณีต่างๆ ที่จัดขึ้นโดยชาวบ้านในชุมชนหรือท้องถิ่น ได้รับความเพลิดเพลินตื่นตาตื่นใจในสุนทรีย์ศิลปะ เพื่อศึกษาวิถีชีวิตและได้รับความรู้ ความเข้าใจ และประสบการณ์ใหม่บนพื้นฐานของความรับผิดชอบต่อและมีจิตสำนึกต่อสิ่งแวดล้อม และมรดกทางวัฒนธรรม โดยคนในท้องถิ่นมีส่วนร่วมต่อการจัดการท่องเที่ยว

1.5) การท่องเที่ยวชมวิถีชีวิตในชนบท (Rural tourism) หมายถึง การเดินทางท่องเที่ยวในหมู่บ้าน ชนบทที่มีลักษณะวิถีชีวิต และผลงานสร้างสรรค์ที่มีเอกลักษณ์พิเศษมีความโดดเด่นเพื่อความเพลิดเพลินได้ความรู้ผลงานสร้างสรรค์และภูมิปัญญาพื้นบ้าน มีความเข้าใจในวัฒนธรรมท้องถิ่น บนพื้นฐานของความรับผิดชอบต่อและมีจิตสำนึกต่อการรักษามรดกทางวัฒนธรรมและคุณค่า ของสภาพแวดล้อม ภายใต้การบริหารจัดการโดยชุมชน

1.6) การท่องเที่ยวอย่างยั่งยืน (Sustainable tourism) [57] หมายถึง การจัดการท่องเที่ยวที่สามารถตอบสนองความต้องการของนักท่องเที่ยวและผู้เป็นเจ้าของท้องถิ่นในปัจจุบัน โดยมีการปกป้องและสงวนรักษาโอกาสต่างๆ ของอนุชนรุ่นหลังด้วยการท่องเที่ยวนี้ รวมถึงการจัดการ

ทรัพยากรเพื่อตอบสนองความจำเป็นทางเศรษฐกิจสังคม และความงามทางสุนทรียภาพ ที่สามารถรักษาเอกลักษณ์ทางวัฒนธรรมและระบบนิเวศน์ด้วย

จากที่กล่าวมาข้างต้นสามารถสรุปได้ว่า “การท่องเที่ยวโดยชุมชน (Community-based tourism: CBT)” เป็นการจัดการท่องเที่ยวทางเลือกที่คำนึงถึงความยั่งยืนของสิ่งแวดล้อม สังคม วัฒนธรรม ให้เกิดการเรียนรู้และสร้างประสบการณ์ใหม่แก่ผู้มาเยือนจากการใช้ประโยชน์ทรัพยากรท้องถิ่นให้คุ้มค่า เพื่อให้คนในชุมชนมีคุณภาพชีวิตที่ดีขึ้น โดยชุมชนเป็นเจ้าของและมีส่วนร่วมในการบริหารจัดการ [1] [8] [24] [58]

2) **เขตพัฒนาการท่องเที่ยวอารยธรรมอีสานใต้** โดยกระทรวงการท่องเที่ยวและกีฬา [51] ได้กำหนดแผนปฏิบัติการพัฒนาการท่องเที่ยวภายในเขตพัฒนาการท่องเที่ยวอารยธรรมอีสานใต้ ซึ่งเป็นหนึ่งในเขตพัฒนาการท่องเที่ยวที่มีศักยภาพการท่องเที่ยวมีความหลากหลาย ทั้งทางธรรมชาติ ประวัติศาสตร์ วัฒนธรรมประเพณีท้องถิ่น และสิ่งปลูกสร้างทันสมัยซึ่งมีความโดดเด่นและมีอัตลักษณ์เฉพาะ รวมถึงชุมชนการท่องเที่ยวภายในเขตพัฒนาการท่องเที่ยวอารยธรรมอีสานใต้ที่มีการดำเนินงานและพัฒนาการมาอย่างต่อเนื่อง บนฐานศักยภาพและบริบทเชิงพื้นที่ที่แตกต่างกันอย่างลงตัว โดยมีเป้าหมายสุดท้ายของการจัดการท่องเที่ยวโดยชุมชนคือ การดูแลทรัพยากรทางธรรมชาติและวัฒนธรรม และเป้าหมายด้านเศรษฐกิจหรือการกระจายรายได้เข้ามาสู่ชุมชนจะเป็นเครื่องหล่อเลี้ยงให้เกิดการยกระดับและพัฒนาขีดความสามารถของการท่องเที่ยวโดยชุมชนได้อย่างมีคุณภาพยิ่งขึ้นและสร้างรายได้จากการท่องเที่ยว ให้เน้นการกระจายประโยชน์อย่างเท่าเทียม มีพื้นที่ครอบคลุมภาคตะวันออกเฉียงเหนือตอนล่าง ประกอบด้วย 5 จังหวัด

2.1) จังหวัดนครราชสีมา หรือที่เรียกกันว่า "โคราช" เป็นเมืองใหญ่บนดินแดนที่ราบสูง ที่อุดมสมบูรณ์ด้วยทรัพยากร และสิ่งอำนวยความสะดวกทุกรูปแบบ ผู้มาเยือนจะได้เพลิดเพลินกับธรรมชาติที่งดงาม จนได้รับการยกย่องให้เป็นมรดกโลก สนุกสนานไปกับกิจกรรมท่องเที่ยวที่หลากหลาย ได้ชื่นชมความยิ่งใหญ่ของอารยธรรมขอมโบราณ เรียนรู้วัฒนธรรมพื้นบ้าน ได้ความรู้ด้านการเกษตร จากการท่องเที่ยวเชิงนิเวศ และเที่ยวเชิงอนุรักษ์ ทั้งยังได้อิ่มอร่อยกับอาหารอีสานต้นตำรับ และเลือกซื้อหาสินค้าเกษตรและหัตถกรรมพื้นบ้าน ในดินแดนที่เปรียบเสมือนเป็นประตูสู่ภาคอีสาน ดังคำขวัญจังหวัดที่ว่า “เมืองหญิงกล้า ผ้าไหมดี หมี่โคราช ปราสาทหิน ดินด่านเกวียน”

2.2) จังหวัดบุรีรัมย์ หมายถึงเมืองแห่งความรื่นรมย์เป็นจังหวัดหนึ่งในภาคอีสาน มีทรัพยากรการท่องเที่ยวที่สำคัญน่าเที่ยว โดยเฉพาะอย่างยิ่งแหล่งโบราณสถานศิลปะแบบขอมโบราณที่มีอยู่มากมายกระจายอยู่ในพื้นที่ทั้งจังหวัด อีกทั้งยังเป็นที่ยังเป็นที่รู้จักในฐานะเมืองเกษตรกรรม

และหัตถกรรมเพราะเป็นแหล่งปลูกข้าวหอมมะลิที่มีคุณภาพดีและเป็นแหล่งทอผ้าไหมที่สวยงาม และมีชื่อเสียงในฐานะเมืองกีฬา (Sport city) นอกจากนี้ด้วยสภาพพื้นที่ตั้งมีอาณาเขตติดต่อกับประเทศเพื่อนบ้านจึงทำให้มีความหลากหลายของเชื้อชาติ ซึ่งก่อให้เกิดความหลากหลายของประเพณีวัฒนธรรมจนทำให้เป็นที่รู้จักโดยทั่วไปดังคำขวัญของจังหวัดที่ว่า “เมืองปราสาทหิน ถิ่นภูเขาไฟ ผ้าไหมสวย รวยวัฒนธรรม เลิศล้ำเมืองกีฬา”

2.3) จังหวัดสุรินทร์ หรือ สะเร็น เป็นแหล่งอารยธรรมเก่าแก่ของชุมชน โบราณมีความเจริญรุ่งเรืองทั้งศิลปกรรม สถาปัตยกรรม วัฒนธรรม ศาสนา เศรษฐกิจ และการปกครอง มีคูเมืองสามชั้น และปราสาทหินที่เก่าแก่ที่สุดและมากที่สุดในประเทศไทย เป็นส่วนหนึ่งของอาณาจักรเจนละ แห่งที่ราบลุ่มตอนกลางของแม่น้ำมูล โดยเป็นศูนย์กลางการติดต่อค้าขายและการประกอบศาสนกิจระหว่างดินแดนภาคตะวันตกและที่ลุ่มต่ำด้านทิศใต้และทิศตะวันออกเฉียงใต้ของเทือกเขาพนมดงรัก มีชนเผ่า “ส่วย กวย กวย” ที่อพยพมาเป็นกลุ่มแรก โดยมีชนกลุ่ม เขมร ลาว ไทย และจีน อพยพตามมา จึงทำให้จังหวัดสุรินทร์มีทั้งศิลปวัฒนธรรมและประเพณีที่หลากหลาย มีความโดดเด่นด้านการท่องเที่ยว อาทิเช่น การเลี้ยงช้าง แหล่งผลิตข้าวหอมมะลิที่มีคุณภาพดีที่สุดในระดับโลก ทอผ้าไหมยกทองที่ที่สวยงามเป็นหนึ่งในแดนสยาม การผลิตเครื่องเงินประจำ และ เมืองเกษตรอินทรีย์ ดังคำขวัญของจังหวัดที่ว่า “สุรินทร์ ถิ่นช้างใหญ่ ผ้าไหมงาม ประคำสวย ร่ำรวยปราสาท ผักกาดหวาน ข้าวสารหอม งามพร้อมวัฒนธรรม”

2.4) จังหวัดศรีสะเกษ มีแหล่งท่องเที่ยวทางธรรมชาติหลายแห่งที่มีลักษณะทางธรรมชาติโดดเด่นและเป็นเอกลักษณ์ในทางธรณีวิทยา วัฒนธรรมขอมโบราณ การค้าชายแดน ซึ่งมีอาณาเขตติดต่อกับประเทศกัมพูชา ตลอดจนแหล่งท่องเที่ยวทางวัฒนธรรมจำนวนมาก ถือเป็นแผ่นดินทองแห่งอีสานใต้ เนื่องจากมีสภาพดินที่อุดมสมบูรณ์ โดยเฉพาะเขตที่ราบลุ่มน้ำมูลและเขตที่ราบลุ่มตอนกลางของจังหวัดในบริเวณที่เรียกว่า ดงภูดินแดง ที่มีลักษณะดินเป็นดินร่วนปนทรายมีความอุดมสมบูรณ์สูง สามารถปลูกพืชเศรษฐกิจสำคัญได้ผลดีเป็นแหล่งพืชสวนและผลไม้อันเป็นผลิตผลทางการเกษตรเพื่อการส่งออกแหล่งใหญ่ที่สำคัญแห่งหนึ่งของประเทศไทย เช่น เงาะ ทุเรียน ภูเขาไฟและสวนผลไม้หลากหลายชนิด ในเดือนมิถุนายนของทุกปีจึงเป็นช่วงเทศกาลของดีศรีสะเกษ เพื่อส่งเสริมการท่องเที่ยวเชิงเกษตรและการท่องเที่ยวเชิงนิเวศของจังหวัด ดังคำขวัญของจังหวัดที่ว่า “หลวงพ่อดำคู่บ้าน ถิ่นฐานปราสาทขอม ข้าวหอมกระเทียมดี มีสวนสมเด็จ เขตดงลำดวน หลากล้วนวัฒนธรรม เลิศล้ำสามัคคี”

2.5) จังหวัดอุบลราชธานี เป็นจังหวัดที่มีศักยภาพด้านการท่องเที่ยวที่โดดเด่น เนื่องจากมีชายแดนติดกับ สปป.ลาว และเป็นเมืองเก่าแก่มีประวัติความเป็นมาที่มากกว่าสองร้อยปี มีสถานที่ท่องเที่ยวหลากหลายและน่าสนใจทั้งด้านประวัติศาสตร์ วัฒนธรรม ประเพณีดั้งเดิมที่ตกทอดจากรุ่นสู่รุ่นมาจากบรรพบุรุษตั้งแต่ยุคขอมโบราณ ผสมผสานเข้ากับวัฒนธรรมไทยถิ่นอีสานปัจจุบันเกิดเป็นเอกลักษณ์เฉพาะตัวและแหล่งท่องเที่ยวทางธรรมชาติที่สวยงามแปลกตา บนที่ราบสูงชัน สลับกับเนินเขาสูง มีแนวหน้าผาริมแม่น้ำโขงหลายแห่ง ตลอดทั้งปีจึงมีการจัดงานเทศกาลเฉลิมฉลองและงานประเพณีที่สำคัญ อาทิเช่น งานแห่เทียนพรรษา งานประเพณีไหลเรือไฟ งานประเพณีมหาสงกรานต์แก่งสะพือ ดังคำขวัญของจังหวัดที่ว่า “อุบลเมืองดอกบัวงาม แม่น้ำสองสี มีปลาแซบหลาย หาดทรายแก่งหิน ถิ่นไทยนักปราชญ์ ทวยราษฎร์ใฝ่ธรรม งามล้ำเทียนพรรษา ผาแต้มก่อนประวัติศาสตร์ ภูทอกภูมิปัญญาท้องถิ่น ดินแดนอนุสาวรีย์คนดีศรีอุบล”

3) **ปัจจัยสำคัญของการท่องเที่ยว (5's Component of tourism)** [25] Sharron ได้กล่าวถึงองค์ประกอบสำคัญ (5A's) ที่แสดงให้เห็นถึงความต้องการที่จำเป็นและส่งผลกระทบต่อความสำเร็จสำหรับการท่องเที่ยว ประกอบด้วย 5 ปัจจัยสำคัญดังนี้

3.1) **สิ่งดึงดูดใจ (Attraction)** [35] [59] เป็นองค์ประกอบที่สำคัญอย่างยิ่ง เป็นสิ่งจูงใจให้เกิดการเดินทางของนักท่องเที่ยว เดินทางไปยังสถานที่ต่าง ๆ อาทิเช่น สิ่งดึงดูดใจเชิงธรรมชาติ ศิลปวัฒนธรรม ชุมชนสัมพันธ์ หรือ สิ่งดึงดูดใจที่มนุษย์สร้างขึ้น

3.2) **กิจกรรม (Activities)** เป็นกิจกรรมที่จัดขึ้นสำหรับนักท่องเที่ยวภายในแหล่งท่องเที่ยวเพื่อเพิ่มความน่าสนใจ ขณะเดียวกันกิจกรรมที่จัดขึ้นควรมีความหลากหลาย เหมาะสม และตรงกับความต้องการของนักท่องเที่ยว อาทิเช่น พายเรือ ปั่นจักรยาน ดูนก เป็นต้น

3.3) **การเข้าถึง (Access)** สภาพเส้นทางคมนาคม การขนส่งและการเดินทางมายังแหล่งท่องเที่ยวที่เหมาะสมและสะดวกสบาย อาทิเช่น สภาพถนน รถโดยสารสาธารณะ เที่ยวบิน หรือ ยานพาหนะที่ให้บริการภายในชุมชน เป็นต้น

3.4) **สิ่งอำนวยความสะดวก (Amenities)** หรือบริการขั้นพื้นฐานที่จำเป็นต่อการจัดการท่องเที่ยวที่มีอยู่ในแหล่งท่องเที่ยวอย่างครบถ้วน ซึ่งนักท่องเที่ยวสามารถใช้ได้อย่างสะดวกสบาย อาทิเช่น ร้านอาหาร อินเทอร์เน็ต โรงพยาบาล หรือ ระบบรักษาความปลอดภัย

3.5) **ที่พัก (Accommodation)** สถานที่ท่องเที่ยวควรมีจำนวนที่พักที่เพียงพอ มีความเหมาะสม สะดวกและปลอดภัยในการเดินทาง พร้อมทั้งมีความหลากหลายด้านราคาและบริการ อาทิเช่น โรงแรม รีสอร์ท บังกะโล โฮมสเตย์ ฯลฯ

4) พฤติกรรมของนักท่องเที่ยว (Tourist behaviours) [26] ใช้ช่วงเวลาและกิจกรรมที่เกิดขึ้นอธิบายถึงแนวคิดพฤติกรรมนักท่องเที่ยวโดยแบ่งพฤติกรรมของนักท่องเที่ยวออกเป็น 5 ช่วงเวลาที่สำคัญ ประกอบด้วย

- 4.1) การสร้างความคาดหวังก่อนการเดินทางท่องเที่ยว (An anticipation or pre-purchase)
- 4.2) เมื่อนักท่องเที่ยวเดินทางถึงสถานที่ท่องเที่ยว (A travel to the site segment)
- 4.3) การได้รับประสบการณ์ระหว่างการเดินทางท่องเที่ยว (An on-site experience)
- 4.4) การเดินทางกลับ (A return travel component) และ
- 4.5) นักท่องเที่ยวจะระลึกถึงประสบการณ์ที่ได้รับจากการท่องเที่ยว (An extended recall and recollection) ที่ให้ความสำคัญต่อการมีปฏิสัมพันธ์ระหว่างนักท่องเที่ยวกับบุคคลอื่น

จากแนวคิดข้างต้นสามารถอธิบายแนวคิด พฤติกรรมนักท่องเที่ยว หมายถึง ความรู้สึกนึกคิดและการแสดงออกในด้านต่าง ๆ ของนักท่องเที่ยว ทั้งก่อนการเดินทางท่องเที่ยว (Pre-trip behaviours) ในระหว่างการเดินทางท่องเที่ยว (Behaviours during the trip) และภายหลังจากการเดินทางท่องเที่ยว (Post-trip behaviours) ตลอดจนปัจจัยต่าง ๆ ที่มีอิทธิพลต่อความรู้สึกนึกคิดและการแสดงออก (Factors influencing tourist behaviours)

5) แรงจูงใจทางการท่องเที่ยว (Tourist motivation) [27] Šimková และคณะ [27] นำเสนอแนวคิด 5 ลำดับขั้นความต้องการของ Maslow (The 5 levels of Maslow's hierarchy of needs) โดยอธิบายไว้อย่างน่าสนใจเพื่อที่จะแสดงให้เห็นว่า มนุษย์โดยทั่วไปมักจะตระหนักถึงความต้องการทางด้านร่างกายเป็นอันดับแรกและเมื่อได้รับการตอบสนองแล้ว เขาก็จะตระหนักถึงความต้องการในขั้นต่อไปตามลำดับดังนี้

- 1) ความต้องการทางร่างกาย (Physiological needs) หมายถึง ความต้องการพื้นฐานของร่างกายที่ทำให้มนุษย์สามารถดำรงชีวิตอยู่ได้อย่างเป็นสุข อาทิเช่น ปัจจัย 4 หรือ ความต้องการพักผ่อน
- 2) ความต้องการความมั่นคงปลอดภัยในชีวิตและทรัพย์สิน (Safety and security needs) หมายถึง ความต้องการความปลอดภัย การปกป้องคุ้มครองในชีวิตและทรัพย์สิน
- 3) ความต้องการความรักและการเป็นส่วนหนึ่งของสังคม (Love and belonging needs) เป็นความต้องการมีปฏิสัมพันธ์ที่ดีและอยู่ร่วมกับผู้อื่นในสังคมอย่างสงบสุข ต้องการความรักและได้รับการยอมรับจากสังคม

4) ความต้องการได้รับการยกย่อง หรือนับถือจากบุคคลอื่น (Self-esteem Needs) เป็นความต้องการด้านชื่อเสียง และให้ผู้อื่นเห็นคุณค่าและยกย่องในความสามารถของตนเอง

5) ความต้องการความสำเร็จและการเข้าใจตัวเองอย่างแท้จริง (Self-actualization needs) เป็นความต้องการที่จะรู้จักตนเองตามสภาพ คุณลักษณะและศักยภาพที่แท้จริงโดยไม่บิดเบือน

Horner [28] ได้เสนอแนวคิดว่าด้วยแรงจูงใจของนักท่องเที่ยวที่หลากหลายซึ่งสามารถแบ่งออกได้เป็น 6 มิติสำคัญด้วยกัน ได้แก่

1) แรงจูงใจทางด้านกายภาพ (Physical motivations) อาทิเช่น ความต้องการพักผ่อน ต้องการให้ผิวเป็นสีแทน ต้องการออกกำลังกาย มีสุขภาพที่ดีขึ้น และความต้องการด้านเพศ

2) แรงจูงใจทางด้านวัฒนธรรม (Cultural motivations) อาทิเช่น ความต้องการชมเมือง และความต้องการได้สัมผัสหรือเรียนรู้ในวัฒนธรรมอื่น ๆ

3) แรงจูงใจทางด้านอารมณ์ (Emotional motivations) อาทิเช่น ความต้องการได้รับความคุ้นเคยในอดีต ต้องการได้รับความรัก ต้องการผจญภัย ต้องการจากชีวิตที่จำเจ ต้องการได้อยู่ในโลกแห่งความเพ้อฝัน และความต้องการเติมเต็มทางจิตวิญญาณหรือความสุข

4) แรงจูงใจทางด้านสถานะ (Status motivations) อาทิเช่น ความต้องการเป็นบุคคลที่สำคัญ ต้องการเป็นผู้มีอำนาจ ต้องการได้รับข้อเสนอที่ดี และความต้องการที่จะใช้จ่ายฟุ้งเฟ้อ

5) แรงจูงใจส่วนบุคคล (Personal motivations) อาทิเช่น ความต้องการไปเยี่ยมเยียนเพื่อน และญาติ ต้องการผูกมิตรกับคนอื่น ต้องการสร้างความพึงพอใจให้กับผู้อื่น และความต้องการได้ใช้ของที่มีราคาประหยัด

6) แรงจูงใจในการพัฒนาตนเอง (Personal development motivations) อาทิเช่น ความต้องการได้รับความรู้เพิ่มเติม และความต้องการได้เรียนรู้ทักษะใหม่ๆ

Šimková และคณะ [27] ประยุกต์ใช้แนวคิดลำดับขั้นความต้องการของ Maslow กับแนวคิดแรงจูงใจทางการท่องเที่ยวของ [28] พอสรุปได้ว่า แรงจูงใจทางการท่องเที่ยว (Tourist motivation) หมายถึง ความต้องการหรือปัจจัยที่ดึงดูดให้นักท่องเที่ยวเข้าไปในแหล่งท่องเที่ยว นั้น ๆ ด้วยวิธีการใดวิธีการหนึ่งบนพื้นฐานความต้องการในมิติที่หลากหลาย เพื่อช่วยให้นักท่องเที่ยวตัดสินใจได้ง่ายขึ้นในการเลือกจุดหมายปลายทาง

6) การแสวงหาความรู้ของนักท่องเที่ยว (Tourist information-seeking behaviours)

Ellis [38] ได้นำเสนอแนวคิดพฤติกรรมแสวงหาสารสนเทศ โดยนำแนวคิดนี้มาประยุกต์ใช้สำหรับการค้นคืนสารสนเทศในงานวิจัยครั้งนี้ ซึ่งพฤติกรรมแสวงหาสารสนเทศนั้น

ยังอยู่ในความสนใจของผู้ใช้อย่างสม่ำเสมอ จะเกิดทักษะในการแสวงหาสารสนเทศด้านการคิดวิเคราะห์ และมีความสามารถในการกำหนดแหล่งและเข้าถึงสารสนเทศที่ต้องการได้รวดเร็ว และได้รับสารสนเทศที่ตนเองต้องการได้ ทักษะการคิดวิเคราะห์ เพื่อการกำหนดแหล่งและเข้าถึงสารสนเทศที่ต้องการนี้ หากได้มีการพัฒนาอย่างสม่ำเสมอ ส่งผลให้มีการพัฒนาทักษะจนเกิดการรู้สารสนเทศ ประกอบด้วย 6 ขั้นตอนสำคัญ ดังนี้

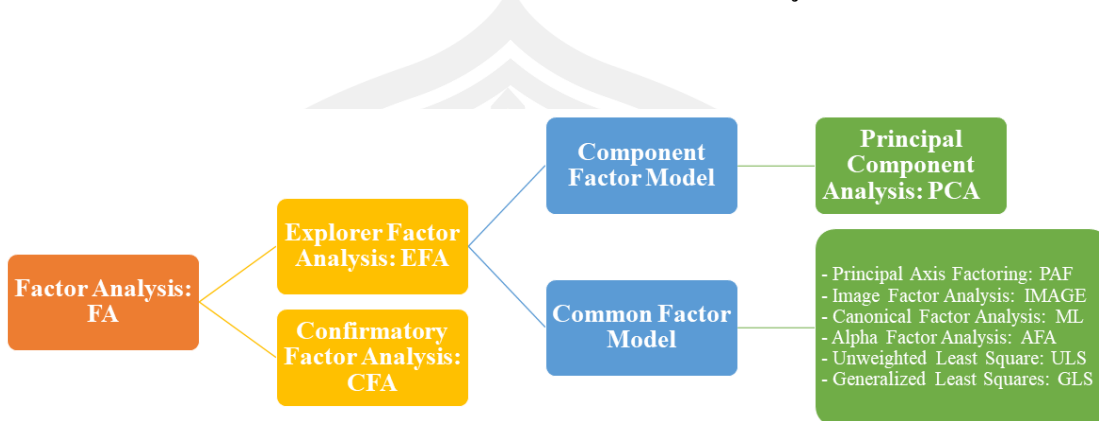
- 1) Starting เป็นการเริ่มต้นกระบวนการแสวงหาสารสนเทศในประเด็นที่สนใจ
- 2) Chaining เป็นการเชื่อมโยงสารสนเทศจากการเริ่มต้นเพื่อเข้าถึงหรือค้นหาสารสนเทศ
- 3) Browsing เป็นการกลั่นกรองสารสนเทศจากการค้นหาอย่างมีหลักการ
- 4) Differentiating เป็นการแยกแยะสารสนเทศที่ค้นหา โดยพิจารณาความถูกต้องและตรงกับความต้องการใช้งาน
- 5) Monitoring เป็นการตรวจตราสารสนเทศ ให้เพียงพอกับความต้องการ
- 6) Extracting เป็นการดึงสารสนเทศที่ต้องการมาเพื่อจัดการในรูปแบบที่ต้องการ

3.2.2 การวิเคราะห์ปัจจัย (Factor analysis: FA)

การวิเคราะห์ปัจจัย [60] เป็นเทคนิควิธีทางสถิติสำหรับการรวมกลุ่มตัวแปรที่มีความสัมพันธ์กันไว้ในกลุ่มเดียวกัน ซึ่งตัวแปรภายในองค์ประกอบเดียวกันจะมีความสัมพันธ์กันสูง ส่วนตัวแปรที่ต่างองค์ประกอบจะสัมพันธ์กันน้อยหรือไม่มี โดยความสัมพันธ์เป็นไปได้ทั้งทางบวกและทางลบ [61] การวิเคราะห์ปัจจัยมีวัตถุประสงค์ 2 ประการหลักคือ 1) ศึกษาว่าองค์ประกอบรวมที่จะสามารถอธิบายความสัมพันธ์ร่วมกันระหว่างตัวแปรต่าง ๆ โดยที่จำนวนองค์ประกอบรวมที่หาได้จะมีจำนวนน้อยกว่าจำนวนตัวแปรนั้น หรือเรียกว่า การวิเคราะห์ปัจจัยเชิงสำรวจ

1) การวิเคราะห์ปัจจัยเชิงสำรวจ เป็นการศึกษาคำสัมพันธ์ระหว่างตัวแปรด้วยหลักการทางสถิติ โดยไม่ได้กำหนดอะไรไว้ล่วงหน้า และ 2) ทดสอบสมมุติฐานในโครงสร้างขององค์ประกอบ หรืออัตราความสัมพันธ์ระหว่างตัวแปรและองค์ประกอบ หรือเรียกว่า การวิเคราะห์ปัจจัยเชิงยืนยัน (Confirmatory factor analysis: CFA) เพื่อยืนยันความสัมพันธ์ของตัวแปร ซึ่งวิธีการดังกล่าวปัจจัยต่างๆจะถูกกำหนดไว้ล่วงหน้า โดยข้อตกลงเบื้องต้นต้องเป็นตัวแปรที่มีค่าต่อเนื่อง หรือ มีค่าในมาตราระดับช่วง (Interval scale) และ มาตราอัตราส่วน (Ratio scale) ซึ่งรูปแบบความสัมพันธ์ระหว่างองค์ประกอบและตัวแปรต้องเป็นแบบเชิงเส้น (Linear) ขณะเดียวกันจำนวนตัวแปรที่คัดเลือกมาควรมีมากกว่า 30 ตัวแปร และกลุ่มตัวอย่างที่ใช้ควรมากกว่าจำนวนตัวแปรอย่างน้อย 5-10 เท่า การวิเคราะห์องค์ประกอบเป็นวิธีการที่ได้รับความนิยมใช้สำหรับลดจำนวนตัวแปร

ให้เหลือน้อยลงเช่นเดียวกับการวิเคราะห์องค์ประกอบหลัก สามารถอธิบายข้อจำกัดและข้อดกเบื้องต้น และ แบ่งประเภทการวิเคราะห์องค์ประกอบได้ 2 ประเภท ดังรูปที่ 8-9



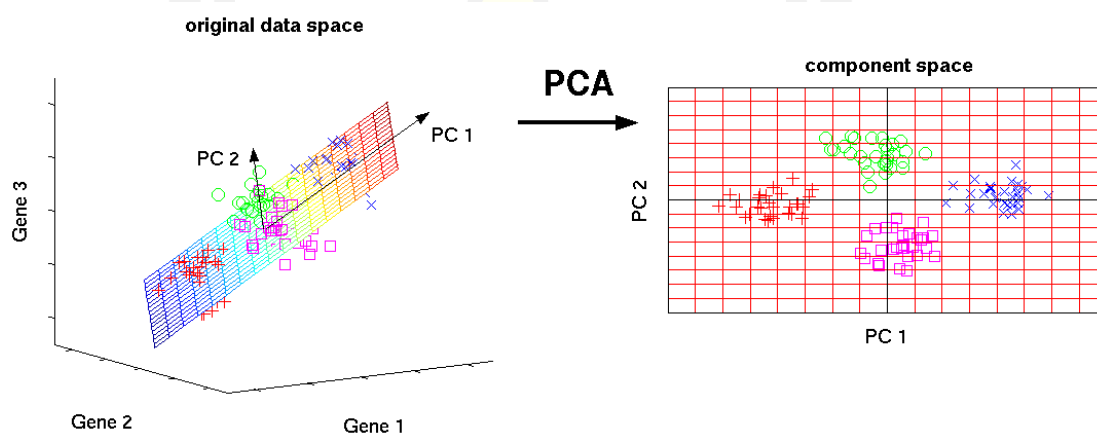
รูปที่ 8 รูปแบบการวิเคราะห์องค์ประกอบ



รูปที่ 9 ข้อจำกัดและข้อดกเบื้องต้นของการใช้สถิติวิเคราะห์องค์ประกอบ

2) การวิเคราะห์องค์ประกอบหลัก [62] เป็นเทคนิคการวิเคราะห์ปัจจัยที่ได้รับความนิยมใช้สำหรับวิเคราะห์ข้อมูลหลายตัวแปร เพื่อหาความสัมพันธ์ของตัวแปรเหล่านั้นส่งผลทำให้

เกิดการลดขนาด Matrix และลดขนาดของคุณลักษณะ (Features) ให้มีขนาดเล็กลง ขณะเดียวกัน การสร้างโมเดลใช้เวลาน้อยลงด้วย [63] โดยการทำงานของวิเคราะห์องค์ประกอบหลัก จะทำการสร้างตัวแปรที่เรียกว่า Component โดยแต่ละ Component จะไม่มีความสัมพันธ์กันเลย Component ตัวแรกจะมีค่า Variance สูงที่สุดและอธิบายข้อมูลได้มากที่สุด ขณะที่ตัวถัดไปจะมี Variance ลดลงตามลำดับ โดย Component ที่เหมาะสมและสามารถอธิบายข้อมูลได้ครอบคลุม Variance ประมาณ 80–90% โดยพิจารณารายละเอียดจากปัจจัยทั้งหมด สร้างการรวมกันของตัวแปรแบบเชิงเส้น (Linear combination) แสดงดังรูปที่ 10



รูปที่ 10 การวิเคราะห์องค์ประกอบหลัก [62]

3.3 ระเบียบวิธีวิจัย

งานวิจัยในครั้งนี้ได้ดำเนินการศึกษารวบรวมข้อมูล แนวคิดและทฤษฎีที่เกี่ยวข้อง นำสู่การวิเคราะห์ข้อมูลด้านการท่องเที่ยว (Tourism domain) โดยชุมชนอีสานใต้ สำหรับวัตถุประสงค์ข้อที่ 1 โดยนำเสนอกระบวนการตามระเบียบวิธีวิจัยได้ 5 ประเด็นสำคัญดังนี้

3.3.1 คำถามวิจัย (Question research)

จุดเริ่มต้นของคำถามวิจัยในวัตถุประสงค์นี้ คือผู้วิจัยยังไม่ทราบถึงโครงสร้างความสัมพันธ์ของตัวแปรที่มีอยู่และยังขาดความรู้ความเข้าใจในประเด็นที่จะศึกษาดีพอ โดยสร้างเครื่องมือเก็บรวบรวมข้อมูลด้วยแบบสอบถาม จำนวน 50 รายการคำถาม นำมาวิเคราะห์ข้อมูลหลายตัวแปรเพื่อหาความสัมพันธ์และสร้างการรวมตัวกันของตัวแปร ผลลัพธ์ที่ได้คือองค์ประกอบใหม่ที่มีความสัมพันธ์กันจากตัวแปรทั้งหมด เพื่อนำผลลัพธ์ที่ได้ไปเป็นฐานในการสร้างกรอบการ

ทำงานการหาความสัมพันธ์คำถามคำตอบอัตโนมัติ นำสู่คำถามวิจัยที่แรกคือ “องค์ประกอบของกลุ่มรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ควรเป็นอย่างไร?”

3.3.2 วัตถุประสงค์การวิจัย (Objective)

เพื่อวิเคราะห์องค์ประกอบของกลุ่มรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้

3.3.3 กรอบแนวคิดการวิจัย (Research framework)

ดำเนินการศึกษารวบรวมข้อมูล แนวคิดและทฤษฎีที่เกี่ยวข้อง นำสู่การวิเคราะห์ข้อมูลด้านการท่องเที่ยว (Tourism domain) โดยชุมชนอีสานใต้ 4 จังหวัด ประกอบด้วย นครราชสีมา บุรีรัมย์ สุรินทร์ และ ศรีสะเกษ โดยวิธีการการคำนวณค่าทางสถิติด้วยการวิเคราะห์องค์ประกอบเชิงสำรวจ เพื่อให้ได้ความสัมพันธ์ของโครงสร้างของตัวแปร โดยเลือกวิธีการวิเคราะห์องค์ประกอบหลัก ที่สามารถอธิบายความผันแปรของชุดข้อมูลได้มากที่สุด [22] ดำเนินการศึกษานโยบายและทฤษฎีที่เกี่ยวข้อง ประกอบไปด้วย 1) 5A ปัจจัยสำคัญของการท่องเที่ยว (5A's Component of tourism) [25] 2) พฤติกรรมนักท่องเที่ยว (Tourist behaviours) [26] 3) แรงจูงใจการท่องเที่ยว (Travel motivation) [27] [28] 4) 4P ส่วนผสมทางการตลาด (4P's Marketing mix) และ 5) หุ่นยนต์สนทนา หรือ ระบบถามตอบอัตโนมัติ (Chatbot or QAs) นำสู่การสร้างแบบสอบถามและตรวจสอบความเหมาะสมของแบบสอบถามโดยผู้เชี่ยวชาญ ก่อนการนำไปใช้งานจริง

3.3.4 ขอบเขตด้านประชากรและกลุ่มตัวอย่าง

1) ประชากร ได้แก่

1.1) คณะกรรมการเครือข่ายการท่องเที่ยวโดยชุมชนเขตอีสานใต้ 4 จังหวัด ประกอบด้วย นครราชสีมา บุรีรัมย์ สุรินทร์ และ ศรีสะเกษ

1.2) กลุ่มนักท่องเที่ยวที่สนใจการท่องเที่ยวโดยชุมชน ในพื้นที่ที่บริหารจัดการท่องเที่ยวโดยชุมชน ภายในเขตอีสานใต้ของประเทศไทยโดยไม่ทราบจำนวนที่แน่นอน

2) กลุ่มตัวอย่าง จำนวน 702 คน

2.1) คณะกรรมการเครือข่ายการท่องเที่ยวโดยชุมชนเขตอีสานใต้ 5 จังหวัด 15 ชุมชน จำนวน 15 คน ประกอบด้วย (1) จังหวัดนครราชสีมา จำนวน 4 ชุมชน ประกอบด้วย ชุมชนบ้านมะค่า ชุมชนตะวันตกวัดเดิม ชุมชนบ้านปรางค์ และ ชุมชนบ้านห้วยแคน (2) จังหวัดบุรีรัมย์ จำนวน 2 ชุมชน ประกอบด้วย ชุมชนบ้านโคกเมือง และ ชุมชนบ้านบุ (3) จังหวัดสุรินทร์ จำนวน 5 ชุมชน ประกอบด้วย ชุมชนบ้านช่างปี ชุมชนบ้านปราสาท ชุมชนบ้านโคกสนวน ชุมชนบ้านสะเดียง

และ ชุมชนบ้านสมบูรณ (4) จังหวัดศรีสะเกษ จำนวน 4 ชุมชน ประกอบด้วย ชุมชนบ้านโนนธาตุ ชุมชนบ้านหนองฮะ ชุมชนบ้านกู่ และ ชุมชนบ้านทามจาม

2.2) กลุ่มนักท่องเที่ยงที่สนใจการท่องเที่ยวโดยชุมชนเขตอีสานใต้ ของประเทศไทย โดยไม่ทราบจำนวนที่แน่นอน ผู้วิจัยใช้วิธีการหาจำนวนกลุ่มตัวอย่างในการเก็บข้อมูลแบบไม่ทราบจำนวนประชากร ด้วยวิธีการสุ่มตัวอย่างแบบกำหนดโควตา (Quota sampling) จากนั้นทำการเลือกตัวอย่างแบบบังเอิญ (Accidental sampling) รวมทั้งสิ้นได้ 702 คน

3.3.5 เครื่องมือที่ใช้ในการวิจัย

1) การสร้างเครื่องมือที่ใช้ในงานวิจัย เริ่มจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง และนำผู้สร้างข้อคำถามตามกรอบแนวคิดการวิจัย โดยแบ่งแบบสอบถามออกเป็น 3 ตอน ประกอบด้วย

ตอนที่ 1 ข้อมูลทั่วไปของนักท่องเที่ยว เป็นข้อมูลพื้นฐานของผู้ตอบแบบสอบถามเพื่อใช้ในการอ้างอิงแหล่งข้อมูล จำนวน 6 ข้อคำถาม ได้แก่ เพศ อายุ สถานภาพ การศึกษาสูงสุด รายได้ และอาชีพ ซึ่งคำถามเป็นแบบเลือกตอบซึ่งเลือกได้คำตอบเดียว

ตอนที่ 2 พฤติกรรมของนักท่องเที่ยว เป็นข้อมูลพฤติกรรม ลักษณะและรูปแบบการเดินทางของนักท่องเที่ยว จำนวน 11 ข้อคำถาม ซึ่งคำถามเป็นแบบเลือกตอบแบ่งเป็น 1) เลือกได้คำตอบเดียว และ 2) เลือกได้หลายคำตอบ โดยผู้วิจัยได้ระบุไว้ในข้อคำถามอย่างชัดเจน ตัวอย่างข้อคำถาม เช่น วัตถุประสงค์ การเดินทาง ที่พัก ระยะเวลา และรูปแบบการท่องเที่ยว ฯลฯ

ตอนที่ 3 รายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวอีสานใต้ เป็นข้อมูลการแสดงความคิดเห็นจากนักท่องเที่ยว จำนวน 50 รายการคำถาม (แสดงดังรูปที่ 17) ลักษณะของแบบสอบถามแบบมาตราส่วนประมาณค่า 5 ระดับ [30] ลิเคิร์ต (Likert scale) ซึ่งผู้วิจัยกำหนดน้ำหนักคะแนนการทำแบบสอบถามดังนี้

รายการคำถามมีอิทธิพลในระดับมากที่สุด	ให้น้ำหนักหรือคะแนนเป็น 5
รายการคำถามมีอิทธิพลในระดับมาก	ให้น้ำหนักหรือคะแนนเป็น 4
รายการคำถามมีอิทธิพลในระดับปานกลาง	ให้น้ำหนักหรือคะแนนเป็น 3
รายการคำถามมีอิทธิพลในระดับน้อย	ให้น้ำหนักหรือคะแนนเป็น 2
รายการคำถามมีอิทธิพลในระดับน้อยที่สุด	ให้น้ำหนักหรือคะแนนเป็น 1

2) ตรวจสอบความตรงของเนื้อหา (Content validity) การวิจัยครั้งนี้ดำเนินการตรวจสอบความตรงของเนื้อหาด้วยค่าดัชนีความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์ (Item-objective congruence index: IOC) จากผู้เชี่ยวชาญ จำนวน 5 คน ที่มีประสบการณ์ด้านการท่องเที่ยวมากกว่า

10 ปีขึ้นไป โดยพิจารณาให้คะแนนตามเกณฑ์ แล้วนำผลคะแนนมาวิเคราะห์หาค่าดัชนีความสอดคล้อง [31] แสดงดังสมการ 3.1 โดยพิจารณาเลือกข้อคำถามที่มีค่า IOC ตั้งแต่ 0.50-1.00 ถือว่ามีความเที่ยงตรงตามเนื้อหา ทั้งนี้หากมีข้อเสนอแนะจากผู้เชี่ยวชาญจะต้องนำมาพิจารณาแก้ไขให้สมบูรณ์ก่อนนำไปวัดผล หรือหาค่าความเชื่อมั่นต่อไป

ระดับคะแนนเท่ากับ 1	คือ มั่นใจในความสอดคล้อง
ระดับคะแนนเท่ากับ 0	คือ ไม่มั่นใจในความสอดคล้อง
ระดับคะแนนเท่ากับ -1	คือ มั่นใจในความไม่ความสอดคล้อง

สมการ

$$IOC = \frac{\sum R}{n} \quad (3.1)$$

เมื่อ $\sum R$ คือ ผลคูณของคะแนนกับจำนวนผู้เชี่ยวชาญ
 n คือ จำนวนผู้เชี่ยวชาญ

ผลการวิเคราะห์ค่าดัชนีความสอดคล้อง^๓ ของผู้เชี่ยวชาญ ด้วยแบบสอบถามการวิเคราะห์รายการคำถามที่มีอิทธิพลต่อการตัดสินใจมาท่องเที่ยวโดยชุมชนเขตอีสานใต้ จำนวน 2 ชุด ประกอบด้วย (1) แบบสอบถามสำหรับสอบถามกลุ่มคณะกรรมการเครือข่ายการท่องเที่ยวโดยชุมชนเขตอีสานใต้ ข้อคำถามแต่ละข้อมีค่าดัชนีความสอดคล้อง (IOC) อยู่ระหว่าง 0.53 – 0.93 วิเคราะห์ค่าความเชื่อมั่นด้วยค่าสัมประสิทธิ์อัลฟ่า (Alpha-reliability coefficient) มีค่าเท่ากับ 0.930 และ (2) แบบสอบถามสำหรับกลุ่มนักท่องเที่ยวที่สนใจการท่องเที่ยวโดยชุมชนเขตอีสานใต้ ประกอบด้วยแบบตรวจรายการสำหรับตอนที่ 1 ข้อมูลทั่วไปของนักท่องเที่ยว จำนวน 6 ข้อ และตอนที่ 2 พฤติกรรมของนักท่องเที่ยว 12 ข้อ และสำหรับตอนที่ 3 เป็นแบบมาตราส่วนประมาณค่า 5 ระดับ จำนวน 50 ข้อ โดยแบบสอบถามได้รับการตรวจสอบความตรง (Validity) จากผู้เชี่ยวชาญจำนวน 5 คน โดยแสดงผลการวิเคราะห์ค่าดัชนีความสอดคล้อง^๓ และปรับแก้ตามข้อเสนอแนะของผู้เชี่ยวชาญ แสดงในภาคผนวก

3) การหาความเชื่อมั่นแอลฟาของครอนบาค (Cronbach's alpha reliability coefficient) [32] ของแบบสอบถามด้วยการทดสอบค่าสัมประสิทธิ์แอลฟาของครอนบาค เป็นหนึ่งในวิธีการหาความเชื่อมั่นแบบความคงที่ภายใน (Internal consistency) และความเที่ยงตรงของแบบสอบถาม [33] ยอมรับค่าสัมประสิทธิ์แอลฟา .70 ขึ้นไป การวิเคราะห์ข้อมูลในขั้นตอนนี้เพื่อตรวจสอบความเข้าใจของผู้ตอบแบบสอบถาม (Pilot survey) โดยทดลองเก็บข้อมูลด้วยแบบสอบถาม (N = 50) จำนวน

50 ข้อ/รายการคำถาม จากกลุ่มตัวอย่างนักท่องเที่ยว (FAM trip) ในจังหวัดสุรินทร์ (N = 36) จำนวน 36 คน มีค่าความเชื่อมั่นหรือความเที่ยงตรงอยู่ในเกณฑ์ดีมาก คือ .983 แสดงรูปที่ 11

Case Processing Summary		
	N	%
Cases		
Valid	36	100.0
Excluded ^a	0	.0
Total	36	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.983	.984	50

รูปที่ 11 ผลการวิเคราะห์การหาความเชื่อมั่น Cronbach's alpha

3) **วิธีการเก็บรวบรวมข้อมูล** ดำเนินการเก็บข้อมูลเชิงปริมาณใน 2 ช่องทาง ประกอบด้วย (1) ผู้วิจัยลงพื้นที่เก็บข้อมูลด้วยตนเอง สำหรับกลุ่มคณะกรรมการเครือข่ายการท่องเที่ยวโดยชุมชนเขตอีสานใต้ จำนวน 15 คน (2) เก็บข้อมูลผ่านช่องทางออนไลน์สำหรับกลุ่มนักท่องเที่ยว จำนวน 328 คน รวมผู้ตอบแบบสอบถามทั้งสิ้น 702 คน ซึ่งคำถามเป็นแบบเลือกตอบและแบบมาตราส่วนประมาณค่า 5 ระดับ เมื่อได้ผลลัพธ์จากการตอบแบบสอบถามของนักท่องเที่ยว ผู้วิจัยเลือกใช้ข้อมูลจากผลการตอบแบบสอบถามในตอนต้นที่ 3 รายการคำถามที่มีอิทธิพลต่อการตัดสินใจเดินทางนักท่องเที่ยว โดยเลือกให้ความสำคัญกับน้ำหนักหรือคะแนนในระดับมาก (4) และ มากที่สุด (5) เพียง 2 ระดับเท่านั้น จากนั้นนำข้อมูลดังกล่าวสู่การวิเคราะห์ด้วยโปรแกรมสำเร็จรูปทางสถิติ เพื่อวิเคราะห์องค์ประกอบเชิงสำรวจในขั้นตอนต่อไป ค่าสถิติที่ใช้ประกอบด้วย การวิเคราะห์องค์ประกอบเชิงสำรวจ เพื่อให้ได้ความสัมพันธ์ของโครงสร้างของตัวแปร โดยเลือกวิธีการวิเคราะห์องค์ประกอบที่สามารถอธิบายการผันแปรของข้อมูลได้มากที่สุด และหมุนแกนองค์ประกอบแบบตั้งฉาก (Varimax orthogonal rotation) สำหรับเกณฑ์ที่ใช้ในการตัดสินใจเลือกตัวแปรที่สำคัญจากองค์ประกอบต่าง ๆ ที่เกิดขึ้นจากการหมุนแกนต้องมีค่าสัมประสิทธิ์ (Factor score coefficients) ตั้งแต่ 0.5 ขึ้นไป

3.4 ผลการทดลอง

การวิเคราะห์องค์ประกอบเชิงสำรวจ เป็นส่วนหนึ่งของเทคนิคการวิเคราะห์องค์ประกอบ (Factor analysis) เพื่อค้นหาความสัมพันธ์โครงสร้างของตัวแปร โดยสามารถลดตัวแปรให้อยู่ในกลุ่มที่มีความสัมพันธ์กันของรายการคำถาม ซึ่งเป็นเครื่องมือที่สำคัญในงานวิจัยครั้งนี้ โดยได้นำวิธีการวิเคราะห์องค์ประกอบหลัก โดยใช้วิธีการหมุนแกนองค์ประกอบแบบตั้งฉาก (Varimax) มีผลทำให้ตัวแปรที่ไม่มีความสัมพันธ์กับกลุ่มหรือมีความสัมพันธ์น้อยจะถูกสกัดออกหรือให้เหลือน้อยที่สุด โดยสังเกตความสัมพันธ์ระหว่างตัวแปรแต่ละตัวกับองค์ประกอบ ได้จากค่าน้ำหนักองค์ประกอบ (Factor loading) จะมีพิสัยระหว่าง -1 ถึง 1

KMO and Bartlett's Test	
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	.982
Approx. Chi-Square	25536.040
Bartlett's Test of Sphericity	df
	1225
	Sig.
	.000

รูปที่ 12 ผลลัพธ์ค่าสถิติ KMO และ Bartlett's test

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	35.898	71.797	71.797	35.898	71.797	71.797	14.336	28.672	28.672
2	2.144	4.288	76.085	2.144	4.288	76.085	12.142	24.285	52.957
3	1.342	2.684	78.770	1.342	2.684	78.770	7.567	15.134	68.091
4	1.173	2.346	81.116	1.173	2.346	81.116	6.513	13.025	81.116
5	.793	1.585	82.701						
6	.689	1.377	84.078						
7	.512	1.024	85.103						
8	.459	.917	86.020						
9	.389	.778	86.798						
10	.349	.698	87.496						
11	.344	.688	88.184						
12	.318	.636	88.820						
13	.290	.580	89.399						
14	.270	.539	89.939						

รูปที่ 13 ผลลัพธ์ค่าความแปรปรวนของชุดข้อมูลทั้งหมด

จากรูปที่ 12 แสดงค่า KMO (Kaiser–Meyer–Olkin) เพื่อวัดค่าความเหมาะสม (Adequacy) ของข้อมูลในการวิเคราะห์องค์ประกอบมีค่าเท่ากับ 0.982 โดยมีค่าประมาณการทางสถิติทดสอบ (Chi-Square) เท่ากับ 25536.040 และมีค่า Significant หรือนัยสำคัญทางสถิติ ที่ระดับ 0.000 สามารถสรุปได้ว่าตัวแปรทั้ง 50 ตัวแปรมีความสัมพันธ์กันเป็นอย่างดีมีความเหมาะสมในการใช้เทคนิคการวิเคราะห์องค์ประกอบ ในการทดสอบสมมติฐาน ซึ่งสมมติฐานในการทดสอบคือ

H0: ตัวแปร ไม่มีความสัมพันธ์กัน และ H1: ตัวแปร มีความสัมพันธ์กัน

โดยใช้สถิติทดสอบ (Chi-Square) มีค่า = 25536 df = 1225 และ sig. = 0.000

จาก sig. = 0.000 < 0.05 ดังนั้นปฏิเสธ H0 ยอมรับ H1

จากรูปที่ 13 แสดงตาราง Total Variance explained แสดงค่าการสกัดปัจจัยทั้งหมดจาก 50 ปัจจัย ได้ 4 กลุ่ม (Cluster) โดยพิจารณาจากตาราง Initial Eigenvalue พบว่า มีเพียง 4 ปัจจัย/องค์ประกอบ ได้แก่ ปัจจัยที่ 1 ค่าตัวแปรมีความสัมพันธ์กันสูงสุดเท่ากับ 35.898 รองลงมาคือ ปัจจัยที่ 2 มีค่าเท่ากับ 2.144 ลำดับต่อมาคือ ปัจจัยที่ 3 มีค่าเท่ากับ 1.342 และปัจจัยที่ 4 ค่าตัวแปรมีความสัมพันธ์กันน้อยที่สุดมีค่าเท่ากับ 1.173 ทั้งนี้ทุกปัจจัยต้องมีค่าความสัมพันธ์ระหว่างตัวแปร หรือค่า Eigenvalues มากกว่า 1 โดยสามารถอธิบายความแปรปรวนของชุดข้อมูลทั้งหมด (% of variance) ได้อย่างชัดเจนที่ร้อยละ 81.116

No. of factor	Component 1		Component 2		Component 3		Component 4	
	factor	factor loading	factor	factor loading	factor	factor loading	factor	factor loading
1	Q37	.769	Q15	.730	Q46	.735	Q3	.803
2	Q38	.763	Q16	.728	Q49	.720	Q4	.775
3	Q31	.761	Q19	.697	Q48	.719	Q2	.772
4	Q34	.756	Q14	.695	Q47	.709	Q1	.727
5	Q42	.752	Q9	.690	Q50	.705	Q5	.716
6	Q33	.751	Q13	.687	Q44	.688		
7	Q32	.750	Q18	.686	Q45	.670		
8	Q36	.749	Q17	.668	Q43	.596		
9	Q40	.740	Q10	.663				
10	Q30	.737	Q7	.654				
11	Q39	.731	Q6	.646				
12	Q41	.723	Q8	.637				
13	Q35	.718	Q11	.636				
14	Q29	.717	Q12	.584				
15	Q27	.691	Q24	.578				
16	Q28	.690	Q20	.574				
17	Q26	.598	Q22	.565				
18	Q23	.528	Q21	.556				
19			Q25	.554				

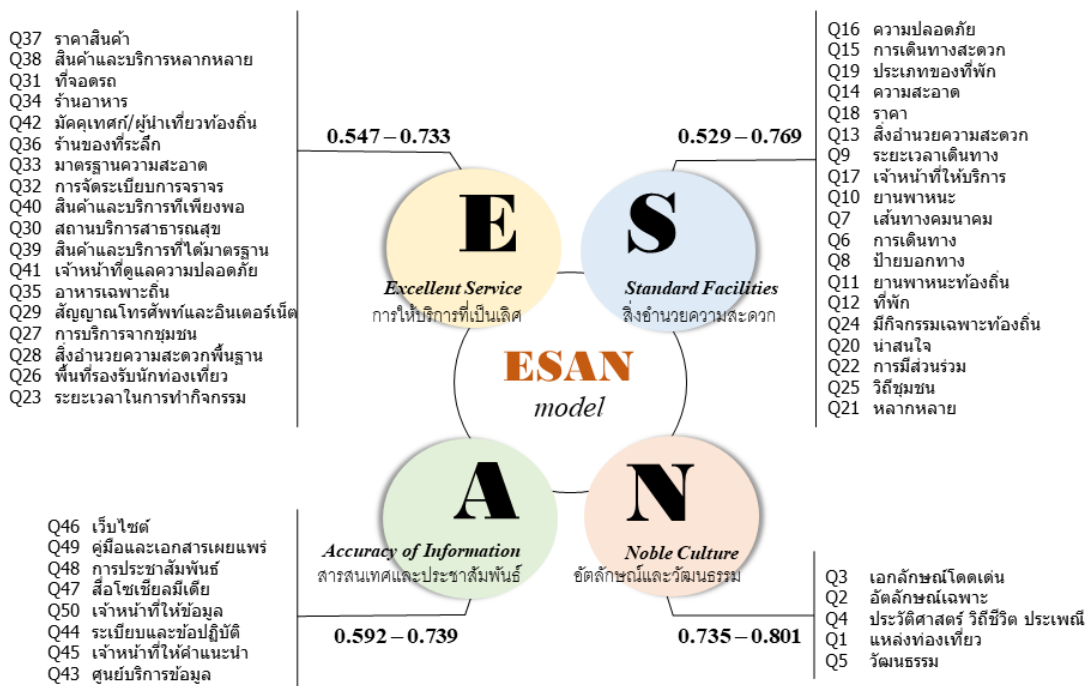
รูปที่ 14 ผลลัพธ์ค่า Factor loading ทั้ง 4 กลุ่ม

จากตารางที่ 14 แสดงค่าน้ำหนักองค์ประกอบ (Factor Loading) เป็นค่าความสัมพันธ์ระหว่างตัวแปรกับองค์ประกอบ โดยกำหนดเกณฑ์ค่าน้ำหนักองค์ประกอบมากกว่า .30 [64] [65]

หากตัวแปรที่ไม่มีความสัมพันธ์กับกลุ่มหรือมีความสัมพันธ์น้อยจะถูกสกัดออกหรือให้เหลือน้อยที่สุด โดยสังเกตความสัมพันธ์ระหว่างตัวแปรแต่ละองค์ประกอบได้จากค่า Factor loading ผลการทดลองเป็นไปตามเกณฑ์ พบว่า องค์ประกอบที่ 1 มี 18 ตัวแปร ให้ค่าอยู่ระหว่าง 0.528-0.769 องค์ประกอบที่ 2 มี 19 ตัวแปร ให้ค่าระหว่าง 0.544-0.730 องค์ประกอบที่ 3 มี 8 ตัวแปร ให้ค่าระหว่าง 0.596-0.735 และองค์ประกอบที่ 4 มี 5 ตัวแปร ให้ค่าระหว่าง 0.716-0.803

3.5 สรุปผลการทดลอง

ผลการทดลองพบว่าองค์ประกอบของรายการคำถามที่มีอิทธิพลต่อการตัดสินใจท่องเที่ยวโดยชุมชนเขตอีสานใต้ เพื่อนำผลลัพธ์จากขั้นตอนนี้ไปสร้างกฎความสัมพันธ์ (Association rule) หรือ ข้อมูลนำเข้าในวัตถุประสงค์ที่สองของงานวิจัย ทั้งหมด 50 ปัจจัย แบ่งได้ 4 กลุ่ม โดยสามารถอธิบายความแปรปรวนของชุดข้อมูลทั้งหมด (% of Variance) ได้ร้อยละ 81.116 โดยผลรวมของน้ำหนักระหว่างองค์ประกอบมีค่าแยกจากกันอย่างชัดเจน ซึ่งในแต่ละองค์ประกอบจะมีตัวแปรไม่เท่ากัน และจัดเรียงตามค่า Factor loading แสดงดังรูปที่ 16 โดยกำหนดชื่อองค์ประกอบใหม่ทั้ง 4 กลุ่ม ดังแสดงดังรูปที่ 15



รูปที่ 15 องค์ประกอบรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวอีสานใต้

กลุ่มที่ 1: ให้ค่าน้ำหนักระหว่างองค์ประกอบอยู่ระหว่าง 0.529-0.769 ประกอบด้วย รายการคำถาม 18 ปัจจัย/องค์ประกอบ ได้แก่ 1) รายการคำถามที่ 37 หรือ Q37: ร้านค้ากำหนดราคาสินค้าและบริการได้อย่างเหมาะสม 2) รายการคำถามที่ 38 หรือ Q38: ร้านค้ามีสินค้าและบริการที่ความหลากหลาย 3) รายการคำถามที่ 31 หรือ Q31: แหล่งท่องเที่ยวมีสถานที่จอดรถที่เหมาะสมและปลอดภัย 4) รายการคำถามที่ 34 หรือ Q34: แหล่งท่องเที่ยวมีร้านอาหารที่สะอาด ได้มาตรฐาน และเพียงพอต่อจำนวนนักท่องเที่ยว 5) รายการคำถามที่ 42 หรือ Q42: แหล่งท่องเที่ยวมีมัคคุเทศก์/ผู้นำเที่ยวท้องถิ่นที่มีความรู้ 6) รายการคำถามที่ 36 หรือ Q36: แหล่งท่องเที่ยวมีร้านจำหน่ายสินค้าชุมชนของฝากและของที่ระลึก 7) รายการคำถามที่ 33 หรือ Q33: แหล่งท่องเที่ยวมีห้องน้ำที่สะอาด มีมาตรฐาน และเพียงพอต่อจำนวนนักท่องเที่ยว 8) รายการคำถามที่ 32 หรือ Q32: แหล่งท่องเที่ยวมีการจัดระเบียบจราจรได้อย่างเหมาะสม 9) รายการคำถามที่ 40 หรือ Q40: ร้านค้ามีสินค้าและบริการที่เพียงพอกับนักท่องเที่ยว 10) รายการคำถามที่ 30 หรือ Q30: แหล่งท่องเที่ยวมีสถานบริการด้านสุขภาพ 11) รายการคำถามที่ 39 หรือ Q39: ร้านค้ามีสินค้าและบริการที่มีคุณภาพได้มาตรฐาน 12) รายการคำถามที่ 41 หรือ Q41: แหล่งท่องเที่ยวมีเจ้าหน้าที่ดูแลความปลอดภัยสำหรับนักท่องเที่ยว 13) รายการคำถามที่ 35 หรือ Q35: มีอาหารที่เป็นเอกลักษณ์เฉพาะของชุมชนบริการแก่นักท่องเที่ยว 14) รายการคำถามที่ 29 หรือ Q29: แหล่งท่องเที่ยวมีสัญญาณโทรศัพท์และสัญญาณอินเทอร์เน็ต ครอบคลุมอย่างทั่วถึง 15) รายการคำถามที่ 27 หรือ Q27: ประชาชนในท้องถิ่นให้การบริการและสนับสนุนด้านการท่องเที่ยวเป็นอย่างดี 16) รายการคำถามที่ 28 หรือ Q28: แหล่งท่องเที่ยวมีสิ่งอำนวยความสะดวกพื้นฐานเพียงพอ 17) รายการคำถามที่ 26 หรือ Q26: ชุมชนมีพื้นที่รองรับนักท่องเที่ยวอย่างเพียงพอ 18) รายการคำถามที่ 23 หรือ Q23: เวลาในการดำเนินกิจกรรมการท่องเที่ยวมีความเหมาะสม โดยกำหนดชื่อใหม่ที่สามารถอธิบายหรือเป็นตัวแทนของกลุ่ม คือ การให้บริการที่เป็นเลิศ (Excellent service หรือ กลุ่ม E)

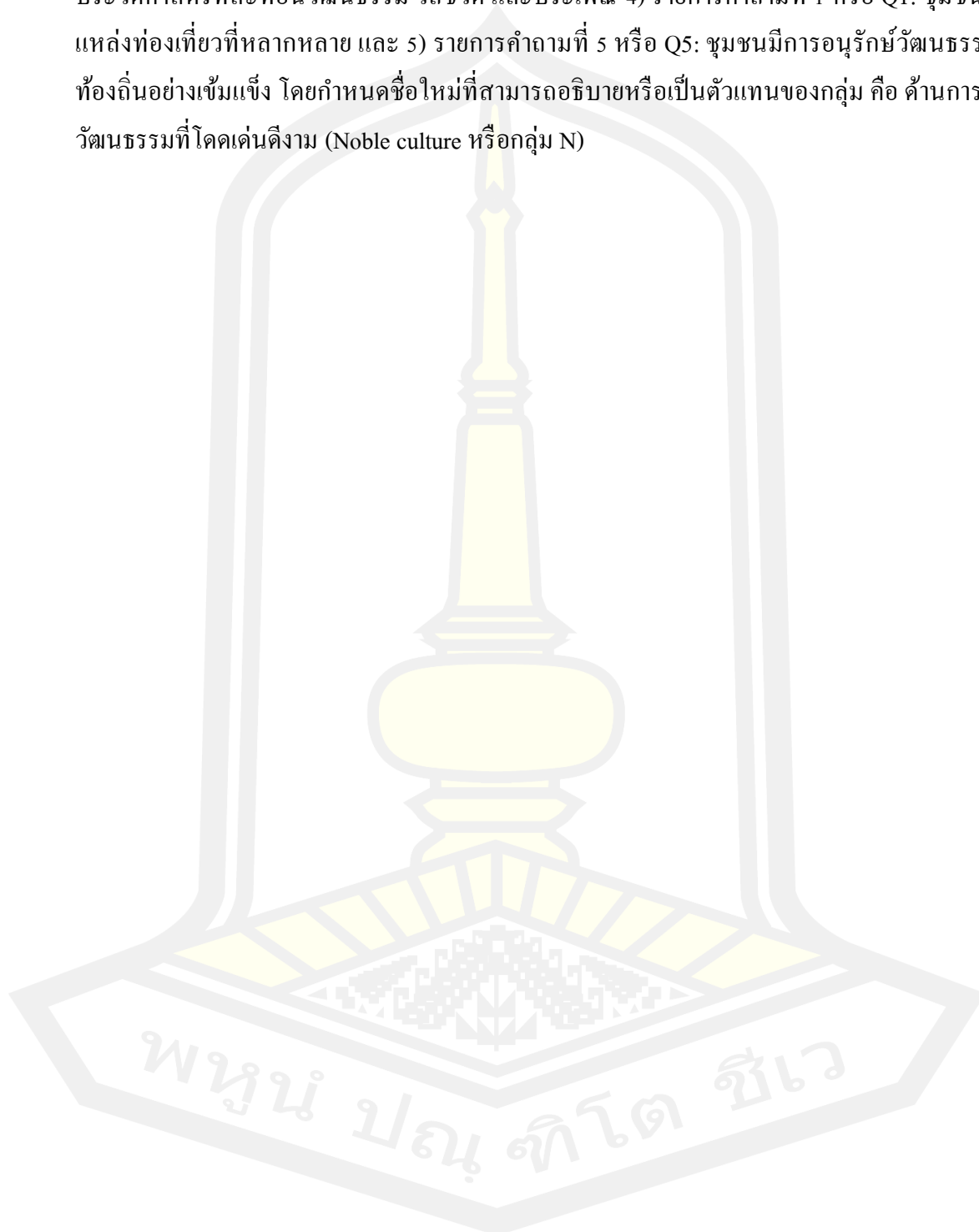
กลุ่มที่ 2 ให้ค่าน้ำหนักระหว่างองค์ประกอบอยู่ระหว่าง 0.547-0.733 ประกอบด้วย รายการคำถาม 19 ปัจจัย/องค์ประกอบ ได้แก่ 1) รายการคำถามที่ 16 หรือ Q16: ที่พักมีความปลอดภัย 2) รายการคำถามที่ 15 หรือ Q15: เข้าถึงที่พักได้อย่างสะดวก 3) รายการคำถามที่ 19 หรือ Q19: ความหลากหลายของประเภทที่พัก 4) รายการคำถามที่ 14 หรือ Q14: ที่พักมีความสะอาด 5) รายการคำถามที่ 18 หรือ Q18: ที่พักมีราคาที่เหมาะสม 6) รายการคำถามที่ 13 หรือ Q13: มีสิ่งอำนวยความสะดวกภายในห้องพัก 7) รายการคำถามที่ 9 หรือ Q9: ระยะเวลาในการเดินทางไปยังแหล่งท่องเที่ยวมีความเหมาะสม 8) รายการคำถามที่ 17 หรือ Q17: ที่พักมีเจ้าหน้าที่/เจ้าของบ้าน

คอยให้ความช่วยเหลือตลอดเวลา 9) รายการคำถามที่ 10 หรือ Q10: มียานพาหนะที่หลากหลายรูปแบบในการเดินทางเข้าถึงแหล่งท่องเที่ยว 10) รายการคำถามที่ 7 หรือ Q7: เส้นทางคมนาคมในการเข้าถึงแหล่งท่องเที่ยวอยู่ในสภาพดี 11) รายการคำถามที่ 6 หรือ Q6: การเดินทางไปยังแหล่งท่องเที่ยวเข้าถึงได้สะดวก 12) รายการคำถามที่ 8 หรือ Q8: มีป้ายบอกเส้นทางในการเข้าถึงแหล่งท่องเที่ยวที่ชัดเจนและมีจำนวนที่เหมาะสม 13) รายการคำถามที่ 11 หรือ Q11: มียานพาหนะสำหรับการเดินทางท่องเที่ยวในพื้นที่ เช่น จักรยาน สามล้อ อีแต่น เกวียน ฯลฯ 14) รายการคำถามที่ 12 หรือ Q12: มีที่พักเพียงพอต่อความต้องการ 15) รายการคำถามที่ 24 หรือ Q24: กิจกรรมการท่องเที่ยวสามารถสะท้อนเอกลักษณ์วัฒนธรรม วิถีชีวิต และภูมิปัญญาท้องถิ่น 16) รายการคำถามที่ 20 หรือ Q20: กิจกรรมการท่องเที่ยวมีความน่าสนใจ 17) รายการคำถามที่ 22 หรือ Q22: มีกิจกรรมการท่องเที่ยวที่เปิดโอกาสให้นักท่องเที่ยวได้มีส่วนร่วม 18) รายการคำถามที่ 25 หรือ Q25: กิจกรรมการท่องเที่ยวสร้างความรู้ความเข้าใจในเอกลักษณ์ของชุมชน และ 19) รายการคำถามที่ 21 หรือ Q21: กิจกรรมการท่องเที่ยวมีความหลากหลาย โดยกำหนดชื่อใหม่ที่สามารถอธิบายหรือเป็นตัวแทนของกลุ่ม คือ ด้านสิ่งอำนวยความสะดวกที่มีมาตรฐาน (Standard facilities หรือ กลุ่ม S)

กลุ่มที่ 3 ให้ค่าน้ำหนักระหว่างองค์ประกอบอยู่ระหว่าง 0.592-0.739 ประกอบด้วย รายการคำถาม 8 ปัจจัย/องค์ประกอบ ได้แก่ 1) รายการคำถามที่ 46 หรือ Q46: แหล่งท่องเที่ยวมีระบบสารสนเทศและเว็บไซต์ให้ข้อมูลกับนักท่องเที่ยว 2) รายการคำถามที่ 49 หรือ Q49: มีคู่มือหรือเอกสารเผยแพร่การท่องเที่ยว และมีรายละเอียดต่าง ๆ ครบถ้วน 3) รายการคำถามที่ 48 หรือ Q48: มีการประชาสัมพันธ์กิจกรรมการท่องเที่ยวผ่าน Social media อย่างต่อเนื่อง 4) รายการคำถามที่ 47 หรือ Q47: แหล่งท่องเที่ยวมีช่องทางการสื่อสารกับนักท่องเที่ยว เช่น Fan page, Facebook, Line 5) รายการคำถามที่ 50 หรือ Q50: มัคคุเทศก์/ผู้นำเที่ยวท้องถิ่นให้บริการข้อมูลที่ดีและมีคุณภาพ 6) รายการคำถามที่ 44 หรือ Q44: แหล่งท่องเที่ยวแจ้งระเบียบและข้อปฏิบัติให้นักท่องเที่ยวทราบอย่างชัดเจน 7) รายการคำถามที่ 45 หรือ Q45: เจ้าหน้าที่ได้คำแนะนำ การบริการที่ดีและมีคุณภาพ (Service mind) และ 8) รายการคำถามที่ 43 หรือ Q43: แหล่งท่องเที่ยวมีการให้บริการด้านข้อมูลและมีศูนย์ข้อมูลแก่นักท่องเที่ยว โดยกำหนดชื่อใหม่ที่สามารถเป็นตัวแทนของกลุ่ม คือ ด้านข้อมูลข่าวสารที่ถูกต้องและทันสมัย (Accuracy of information หรือกลุ่ม A)

กลุ่มที่ 4 ให้ค่าน้ำหนักระหว่างองค์ประกอบอยู่ระหว่าง 0.735-0.801 ประกอบด้วย รายการคำถาม 5 ปัจจัย/องค์ประกอบ ได้แก่ 1) รายการคำถามที่ 3 หรือ Q3: ชุมชนมีแหล่งท่องเที่ยวที่เป็นเอกลักษณ์ (Uniqueness) โดดเด่น 2) รายการคำถามที่ 2 หรือ A2: ชุมชนมีแหล่งท่องเที่ยวที่

มีอัตลักษณ์ (Identity) เฉพาะท้องถิ่น 3) รายการคำถามที่ 4 หรือ Q4: ชุมชนมีแหล่งท่องเที่ยวทางประวัติศาสตร์ที่สะท้อนวัฒนธรรม วิถีชีวิต และประเพณี 4) รายการคำถามที่ 1 หรือ Q1: ชุมชนมีแหล่งท่องเที่ยวที่หลากหลาย และ 5) รายการคำถามที่ 5 หรือ Q5: ชุมชนมีการอนุรักษ์วัฒนธรรมท้องถิ่นอย่างเข้มแข็ง โดยกำหนดชื่อใหม่ที่สามารถอธิบายหรือเป็นตัวแทนของกลุ่ม คือ ด้านการมีวัฒนธรรมที่โดดเด่นดีงาม (Noble culture หรือกลุ่ม N)



บทที่ 4

การค้นหารูปแบบคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้

การดำเนินการในวัตถุประสงค์ที่สองของการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาเอกสารและผลงานวิจัยที่เกี่ยวข้องกับสาระสำคัญเพิ่มเติมจากวัตถุประสงค์ที่หนึ่ง เพื่อให้เกิดความเข้าใจที่ชัดเจน ตรงประเด็น บนฐานคิดจากคำถามวิจัย “รูปแบบรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ควรเป็นอย่างไร?” โดยมีวัตถุประสงค์เพื่อค้นหารูปแบบคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ โดยผู้วิจัยนำเสนอใน 6 ประเด็นสำคัญ ประกอบด้วย 1) ที่มาและความสำคัญของปัญหา 2) เอกสารและงานวิจัยที่เกี่ยวข้อง 3) ระเบียบวิธีวิจัย 4) ผลการทดลอง 5) สรุปผลการทดลอง และ 6) อภิปรายผล ตามลำดับดังนี้

4.1 ที่มาและความสำคัญของปัญหา

อุตสาหกรรมท่องเที่ยวนับได้ว่าเป็นอุตสาหกรรมสำคัญที่สร้างงานและสร้างอาชีพให้กับประชาชน ตลอดจนสร้างรายได้ที่เป็นเงินตราต่างประเทศมาเป็นระยะเวลายาวนาน เมื่อเปรียบเทียบกับอัตราการเติบโตเฉลี่ยของจำนวนนักท่องเที่ยวทั่วโลกในปี 2560-2561 ของ UNWTO พบว่าสูงกว่าการคาดการณ์ โดยประเทศไทยเป็นประเทศที่ได้รายได้จากนักท่องเที่ยวมากที่สุดในภูมิภาคราว 57.5 พันล้านดอลลาร์สหรัฐ [13] ปัจจุบันพฤติกรรมและความคาดหวังของนักท่องเที่ยวที่ปรับเปลี่ยนไปบนความเจริญก้าวหน้าของเทคโนโลยีสารสนเทศและสื่อโซเชียลมีเดีย ส่งผลต่อพฤติกรรมแสวงหาความรู้และการค้นหาข้อมูลทางการท่องเที่ยว การบริหารจัดการสารสนเทศด้านการท่องเที่ยวสะดวกและรวดเร็วมากยิ่งขึ้น ส่งผลให้นักท่องเที่ยวเลือกเดินทางท่องเที่ยวด้วยตนเองมากขึ้น ซึ่งประเทศไทยมีความจำเป็นต้องปรับตัวต่อสถานการณ์ดังกล่าวอย่างรู้เท่าทัน [21]

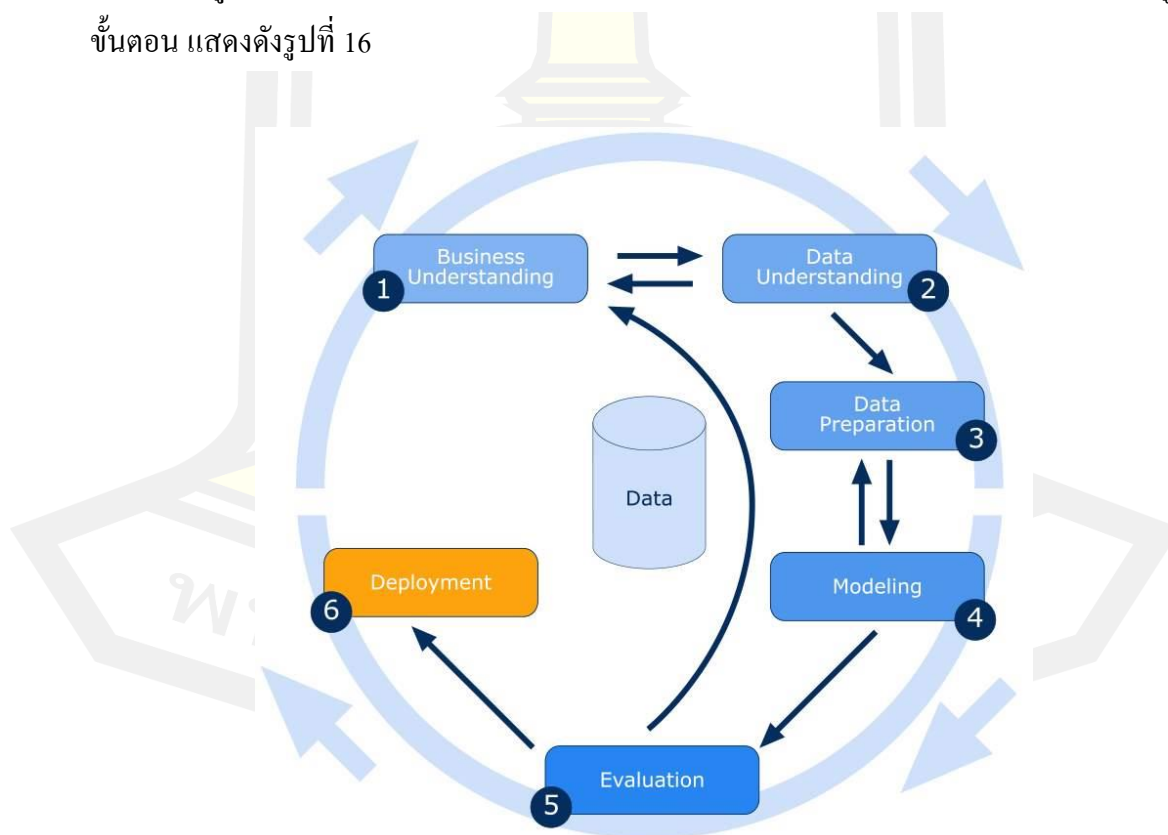
การออกแบบคำถามที่ดีเป็นจุดเริ่มต้นสู่ความสำเร็จของการออกแบบระบบถามตอบอัตโนมัติ [36] [53] นำสู่การได้มาซึ่งคำตอบที่ถูกต้องและตรงกับความต้องการและส่งผลต่อการตัดสินใจ การค้นหาความสัมพันธ์เป็นกระบวนการหนึ่งที่มีเป้าหมายในการหาความเกี่ยวเนื่องและความสัมพันธ์ระหว่างข้อมูลเป็นวิธีการที่มีลักษณะเฉพาะตัวสำหรับแก้ไขปัญหาข้อมูลแบบมีโครงสร้าง [66] การนำกฎความสัมพันธ์ไปใช้งานจริงในปัจจุบัน [49] ได้ถูกนำไปประยุกต์ใช้งานอย่างกว้างขวาง นำสู่การค้นหารูปแบบการสอบถามข้อมูลของนักท่องเที่ยว เพื่อนำไปปรับปรุงการให้บริการระบบถามตอบอัตโนมัติ แสดงให้เห็นรูปแบบการถามคำถาม นำสู่การปรับปรุงกลยุทธ์รูปแบบการให้บริการ [36] เพิ่มประสิทธิภาพในการเข้าถึงลูกค้า รวมทั้งการตอบโต้ระหว่างผู้ซื้อและผู้ให้บริการ ก่อน ระหว่างและหลังรับบริการ เพื่อตอบสนองความต้องการของนักท่องเที่ยว

4.2 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาเอกสารและผลงานวิจัยที่เกี่ยวข้องกับสาระสำคัญเพิ่มเติม จากวัตถุประสงค์ข้อแรก เชื่อมโยงเพื่อนำสู่การค้นหารูปแบบคำถามที่มีความสัมพันธ์ต่อการ ท่องเที่ยวโดยชุมชนอีสานใต้ โดยนำเสนอวรรณกรรมที่เกี่ยวข้องใน 4 ประเด็น ได้แก่ 1) เหมือน ข้อมูล (Data mining: DM) 2) เหมือนข้อความ (Text mining: TM) 3) กฎความสัมพันธ์ (Association rule mining: ARM) 4) วิธีแอฟริอริ (Apriori algorithm) ตามลำดับดังนี้

4.2.1 เหมือนข้อมูล (Data Mining: DM)

เหมือนข้อมูล (Data Mining) [41] เป็นวิธีการที่ดำเนินการกับข้อมูลจำนวนมากเพื่อ ค้นหารูปแบบและความสัมพันธ์ ที่ซ่อนอยู่ในชุดข้อมูล สามารถจัดเก็บในรูปแบบข้อมูลที่สามารถ ดึงข้อมูลสารสนเทศมาใช้ทำให้ค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล ในปัจจุบันการทำเหมือนข้อมูลได้ ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในเชิงพาณิชย์เพื่อช่วยในการตัดสินใจ ด้าน วิทยาศาสตร์และการแพทย์รวมถึงด้านเศรษฐกิจและสังคม กระบวนการมาตรฐานสำหรับการทำ เหมือนข้อมูล เพื่อทำการวิเคราะห์ (CRISP-DM) และนำไปใช้ประโยชน์ในเชิงพาณิชย์ มีอยู่ 6 ขั้นตอน แสดงดังรูปที่ 16



รูปที่ 16 กระบวนการมาตรฐานที่ใช้สำหรับการทำเหมือนข้อมูล (CRISP-DM) [41]

- 1) การทำความเข้าใจธุรกิจ (Business understanding) เป็นขั้นในการทำความเข้าใจธุรกิจ ปัญหาและความต้องการ นำสู่การวิเคราะห์ข้อมูลและวางแผนปฏิบัติการ
- 2) การทำความเข้าใจข้อมูล (Data understanding) เป็นการรวบรวมข้อมูล ทำความเข้าใจ ตรวจสอบคุณภาพ โดยคัดเลือกข้อมูลที่ต้องการนำมาวิเคราะห์
- 3) การเตรียมข้อมูล (Data preparation) เป็นการเปลี่ยนข้อมูลดิบให้เป็นข้อมูลที่สมบูรณ์ นำสู่การสร้างโมเดลในขั้นตอนต่อไป
- 4) การสร้างโมเดล (Modeling) เป็นการเลือกและทดลองสร้าง โมเดลหลาย ๆ แบบ สำหรับแก้ไขปัญหที่ต้องการเพื่อให้ได้โมเดลที่เหมาะสมที่สุดสำหรับการแก้ไขปัญห
- 5) การวัดประสิทธิภาพของโมเดล (Evaluation) เป็นการวัดประสิทธิภาพของโมเดล เพื่อให้เหมาะกับการนำไปใช้งาน ซึ่งวิธีการวัดประสิทธิภาพ โมเดลจะแตกต่างกันออกไป
- 6) การนำโมเดลไปใช้งานจริง (Deployment) เป็นการนำโมเดลที่เหมาะสมที่สุดไปใช้งานจริง เพื่อวิเคราะห์และแก้ปัญหที่ต้องการ

4.2.2 เหมืองข้อความ (Text Mining: TM)

เหมืองข้อความ [67] เป็นกระบวนการสกัดเอาความรู้จากภาษาธรรมชาติและการค้นหา กฎความสัมพันธ์ของข้อความ โดยการวิเคราะห์ชุดข้อมูลส่วนใหญ่มีลักษณะแบบไม่มีโครงสร้าง หรือกึ่งโครงสร้างจำนวนมาก เพื่อให้ได้มาซึ่งสารสนเทศที่มีคุณค่า การทำนายเหตุการณ์หรือจัดกลุ่มของข้อมูลได้อย่างมีประสิทธิภาพ ปัจจุบันนิยมนำมาใช้งานร่วมกับองค์ความรู้หลายด้าน อาทิ เช่น เหมืองข้อมูล (Data mining) การประมวลผลภาษาธรรมชาติ (Natural language processing: NLP) การค้นคืนสารสนเทศ (Information retrieval) ฐานการเรียนรู้เครื่องจักร (Machine learning) แขนงหนึ่งของปัญญาประดิษฐ์ (Artificial intelligence: AI)

4.2.3 กฎความสัมพันธ์ (Association Rule Mining: ARM)

กฎความสัมพันธ์ [41], [66] เป็นอัลกอริทึมแบบ Unsupervised เทคนิคหนึ่งของการทำ Data Mining ที่สำคัญ และสามารถนำไปประยุกต์ใช้ได้จริง มีหลักการทำงานโดยการค้นหาความสัมพันธ์ของข้อมูลจากฐานข้อมูลขนาดใหญ่ที่มีอยู่เพื่อนำไปใช้ในการวิเคราะห์ หรือทำนายปรากฏการณ์ต่าง ๆ ส่วนใหญ่จะเป็นการ “Market basket analysis” ซึ่งผลการวิเคราะห์ที่ได้จากการค้นหาความสัมพันธ์ของข้อมูลจะเป็นคำตอบของปัญหา ในการหาความสัมพันธ์นั้นจะมีขั้นตอนวิธีการหาหลายวิธีด้วยกัน ซึ่งขั้นตอนวิธีที่เป็นที่รู้จักและใช้อย่างแพร่หลายคือ Apriori

4.2.4 เอพริออริ (Apriori algorithm)

เอพริออริ [18], [68] นิยมนำมาค้นหาความสัมพันธ์ที่เกิดขึ้นซ้ำ ๆ จากชุดข้อมูล สำหรับชุดข้อมูลขนาดเล็กจะสามารถค้นหาความสัมพันธ์ได้อย่างรวดเร็วและจะใช้เวลาค้นหาเพิ่มมากขึ้นสำหรับข้อมูลขนาดใหญ่และมีความซับซ้อนมาก โดยพิจารณากฎ $A \rightarrow B$ คือการซื้อสินค้า A แล้วซื้อสินค้า B จะเกิดขึ้นและมีความเชื่อมั่น ได้ที่เปอร์เซ็นต์ ผลลัพธ์ที่ได้จะถูกนำไปใช้แนะนำลูกค้าให้ซื้อสินค้าเพิ่มทั้งแบบ Up-selling และ Cross-selling ใช้สำหรับการปรับกลยุทธ์การขายและพิจารณาการจัดวางสร้างความสะดวกในการเลือกซื้อสินค้า การจัดโปรโมชั่นและเพิ่มยอดขายจำหน่ายให้กับร้านค้าและธุรกิจ ซึ่งการหาความสัมพันธ์ที่แข็งแกร่งต้องผ่านค่า Support คือค่าความน่าจะเป็นของจำนวน Transaction ที่ปรากฏร่วมกันของ A และ B โดย Confidence คือค่าความน่าจะเป็นของจำนวน Transaction ที่ปรากฏร่วมกันของ A และ B หาด้วยความน่าจะเป็นของ Transaction ที่ปรากฏ A อย่างเดียว ดังสมการที่ 4.1-4.2

$$Support(A) = \frac{\text{Number of transaction in which A appears}}{\text{Total number of transactions}} \quad (4.1)$$

$$Confidence(A \rightarrow B) = \frac{\text{support}(A \cup B)}{\text{support}(A)} \quad (4.2)$$

4.3 ระเบียบวิธีวิจัย

งานวิจัยในครั้งนี้ได้ดำเนินการศึกษารวบรวมข้อมูล แนวคิดและทฤษฎีที่เกี่ยวข้อง นำสู่การวิเคราะห์ข้อมูลด้านการท่องเที่ยว (Tourism domain) โดยชุมชนอีสานใต้ สำหรับวัตถุประสงค์ข้อที่ 3 โดยนำเสนอกระบวนการตามระเบียบวิธีวิจัยได้ 5 ประเด็นสำคัญดังนี้

4.3.1 คำถามวิจัย (Question research)

รูปแบบรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ควรเป็นอย่างไร?

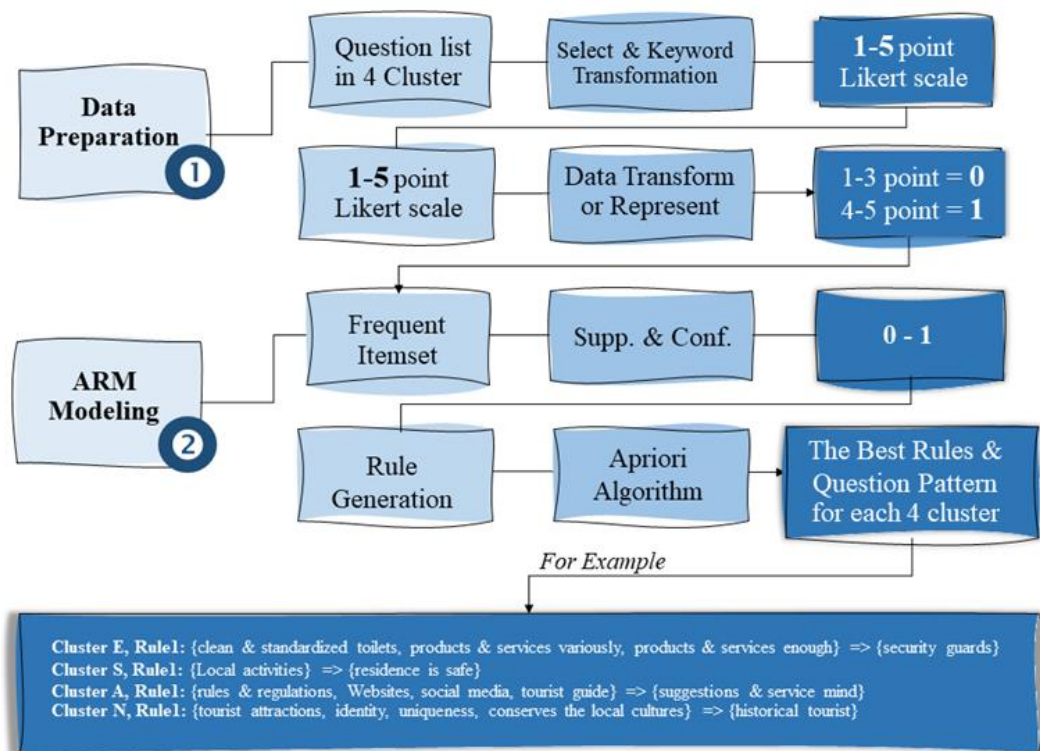
4.3.2 วัตถุประสงค์การวิจัย (Objective)

เพื่อกันหารูปแบบคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้

4.3.3 กรอบแนวคิดการวิจัย (Research framework)

ดำเนินการศึกษาแนวคิดและทฤษฎีที่เกี่ยวข้อง ประกอบไปด้วย ได้แก่ 1) เหมืองข้อมูล 2) เหมืองข้อความ 3) กฎความสัมพันธ์ 4) Apriori algorithm ตามกรอบแนวคิดการวิเคราะห์เหมืองข้อมูล CRISP-DM และการประยุกต์ใช้กฎความสัมพันธ์ ในการวิเคราะห์ความสัมพันธ์รายการ

คำถามที่มีอิทธิพลต่อการตัดสินใจเดินทางของนักท่องเที่ยว ด้วย Apriori algorithm ใน โปรแกรม Weka สำหรับค้นหารูปแบบการสอบถามหรือการแสวงหาข้อมูลของนักท่องเที่ยวที่ปรากฏร่วมกันบ่อย ๆ (Frequent item set) นำเสนอกระบวนการวิจัยดังรูปที่ 17



รูปที่ 17 กรอบการดำเนินงานการวิจัยตามวัตถุประสงค์ที่ 2

4.3.4 การเก็บรวบรวมข้อมูล (Data collection)

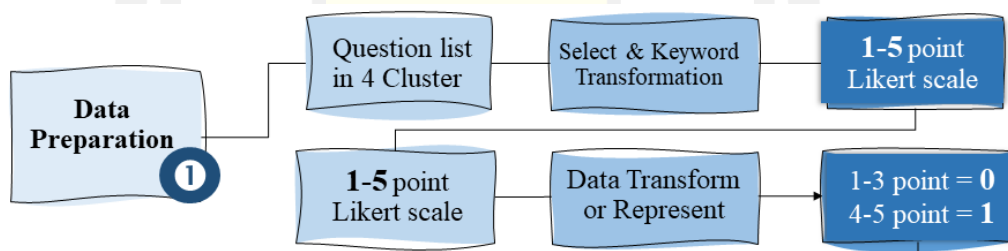
ผู้วิจัยใช้ข้อมูลเดิมจากวัตถุประสงค์ข้อ 1 ที่ดำเนินการเก็บรวบรวมข้อมูล รวมทั้งสิ้น 702 คน/ตัวอย่าง 50 รายการคำถาม ด้วยแบบสอบถามออนไลน์จากกลุ่มตัวอย่างนักท่องเที่ยวของเครือข่ายการท่องเที่ยวโดยชุมชนอีสานใต้ใน 4 จังหวัดเช่นเดิม การวิเคราะห์ข้อมูลในครั้งนี้ ทำความเข้าใจข้อมูล (Data understanding) ตรวจสอบคุณภาพและการคัดเลือกข้อมูลที่จะนำมาวิเคราะห์เป็นสิ่งสำคัญ ผู้วิจัยเลือกใช้ข้อมูลจากผลลัพธ์ในตอนต้นที่ 3 รายการคำถามที่มีอิทธิพลต่อการตัดสินใจเดินทางของนักท่องเที่ยว โดยเลือกให้ความสำคัญกับน้ำหนักหรือคะแนนในระดับมาก (4) และ มากที่สุด (5) เพียง 2 ระดับเท่านั้น

4.3.5 การเตรียมข้อมูล (Data preparation)

ผู้วิจัยเลือกใช้ข้อมูลผลการตอบแบบสอบถามในตอนที่ 3 จำนวน 50 รายการคำถาม จากกลุ่มตัวอย่างผู้ตอบแบบสอบถามรวม 702 คน จัดเก็บและบันทึกข้อมูลลงในโปรแกรมไมโครซอฟท์เอ็กเซล คัดเลือกข้อมูลและกลั่นกรองข้อมูลด้วยการลบข้อมูลในตอนที่ 1 และ 2 ที่ทิ้งไป ซึ่งในขั้นตอนนี้จะเหลือเฉพาะข้อมูลตอนที่ 3 โดยพิจารณาให้มีความสำคัญกับน้ำหนักหรือคะแนนจากการตอบแบบสอบถามเพียง 2 ระดับ คือ ระดับมาก (4) และ มากที่สุด (5) ซึ่งการใช้ Apriori algorithm ใน โปรแกรม Weka มีข้อจำกัดคือ ข้อมูลที่นำมาใช้จะอยู่ในรูปแบบ Nominal หรือ Ordinal เท่านั้น โดยผู้วิจัยเลือกแทนค่าข้อมูลน้ำหนักคะแนนด้วยเลข 0 และ 1

ระดับน้อยที่สุด (1) ระดับน้อย (2) และ ระดับปานกลาง (3)	แทนค่าด้วย	0
ระดับมาก (4) และ ระดับมากที่สุด (5)	แทนค่าด้วย	1

เข้าสู่กระบวนการแปลงรูปข้อมูล (Data transformation) ด้วยการจัดเก็บและบันทึกข้อมูลลงในโปรแกรมไมโครซอฟท์เอ็กเซล ให้อยู่ในรูปแบบไฟล์ .CSV เพื่อให้สามารถนำข้อมูลมาวิเคราะห์รายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ ด้วยโปรแกรม Weka ในขั้นตอนต่อไป แสดงดังรูปที่ 18



รูปที่ 18 ขั้นตอนการเตรียมข้อมูล (Data preparation) 4 กลุ่ม (Cluster)

4.4 ผลการทดลอง

การวิจัยในครั้งนี้ได้นำเข้าข้อมูลจากผลการตอบแบบสอบถามที่ปรากฏร่วมกันบ่อย (Frequent item sets) ของนักท่องเที่ยว จำนวน 702 เรคคอร์ด 50 รายการคำถาม แบ่งการวิเคราะห์ข้อมูลออกเป็น 4 กลุ่ม (Clusters) นำสู่การสร้างกฎความสัมพันธ์และค้นหารูปแบบการแสวงหาความรู้ของนักท่องเที่ยวด้วย Apriori algorithm โดยกำหนดค่าเริ่มต้นในโปรแกรม Weka 3.6 (minSupport = 0.1, minMetric = 0.9, numRules = 100) ดำเนินการวิเคราะห์ข้อมูลที่ละกลุ่ม จำนวน

4 ครั้ง ได้กฎความสัมพันธ์รวมทั้งสิ้น 400 กฎ โดยเลือกนำเสนอตัวอย่างกฎความสัมพันธ์จากการพิจารณาประเด็นสำคัญใน 2 ขั้นตอนดังนี้

1) พิจารณาเลือกกฎความสัมพันธ์ของตัวชี้วัดประสิทธิภาพจากค่า Confidence ที่ยอมรับได้ตั้งแต่ 98% ขึ้นไป

2) พิจารณาเลือกกฎที่มีลักษณะเด่น มีความเฉพาะที่เป็นไปในทิศทางเดียวกันในเชิงประจักษ์และสอดคล้องกับแนวคิด 5 ลำดับขั้นความต้องการของ Maslow โดยตัดกฎที่มีความซ้ำซ้อนทิ้ง [40]

Class E	
rule1: Q33=1 Q38=1 Q40=1 391 ==> Q41=1 386	<conf:(0.99)> lift:(1.5) lev:(0.18) [128] conv:(22.28)
rule2: Q37=1 Q38=1 Q40=1 397 ==> Q41=1 391	<conf:(0.98)> lift:(1.5) lev:(0.18) [129] conv:(19.39)
rule3: Q34=1 Q38=1 Q40=1 396 ==> Q41=1 390	<conf:(0.98)> lift:(1.5) lev:(0.18) [129] conv:(19.34)
rule4: Q35=1 Q38=1 Q40=1 396 ==> Q41=1 390	<conf:(0.98)> lift:(1.5) lev:(0.18) [129] conv:(19.34)
Class S	
rule1: Q24=1 702 ==> Q16=1 702	<conf:(1)> lift:(1) lev:(0) [0] conv:(0)
rule2: Q16=1 702 ==> Q24=1 702	<conf:(1)> lift:(1) lev:(0) [0] conv:(0)
rule3: Q25=1 476 ==> Q16=1 476	<conf:(1)> lift:(1) lev:(0) [0] conv:(0)
rule4: Q25=1 476 ==> Q24=1 476	<conf:(1)> lift:(1) lev:(0) [0] conv:(0)
Class A	
Rule1: Q44=1 Q46=1 Q48=1 Q49=1 354 ==> Q45=1 352	<conf:(0.99)> lift:(1.58) lev:(0.18) [129] conv:(43.7)
Rule2: Q44=1 Q46=1 Q47=1 Q49=1 353 ==> Q45=1 351	<conf:(0.99)> lift:(1.58) lev:(0.18) [128] conv:(43.58)
Rule3: Q44=1 Q46=1 Q47=1 Q48=1 363 ==> Q45=1 360	<conf:(0.99)> lift:(1.58) lev:(0.19) [131] conv:(33.61)
Class N	
Rule1: Q1=1 Q2=1 Q3=1 Q5=1 293 ==> Q4=1 292	<conf:(1)> lift:(1.7) lev:(0.17) [120] conv:(60.52)
Rule2: Q2=1 Q3=1 Q5=1 323 ==> Q4=1 321	<conf:(0.99)> lift:(1.69) lev:(0.19) [131] conv:(44.48)
Rule3: Q1=1 Q2=1 Q5=1 299 ==> Q4=1 297	<conf:(0.99)> lift:(1.69) lev:(0.17) [121] conv:(41.17)

รูปที่ 19 ตัวอย่างผลลัพธ์การสร้างกฎความสัมพันธ์ด้วยโปรแกรม Weka

จากรูปที่ 19 แสดงตัวอย่างผลการวิเคราะห์กฎความสัมพันธ์ของรายการคำถามด้วยอัลกอริทึม Apriori ผู้วิจัยได้พิจารณาเลือกกฎความสัมพันธ์ที่ดีที่สุดโดยเลือกกำหนดกฎความสัมพันธ์ที่ 100 กฎ (numRules = 100) จากตัวชี้วัดประสิทธิภาพจากค่า Confidence ที่ยอมรับได้ตั้งแต่ 98% ขึ้นไป สามารถสร้างกฎที่ดีที่สุดและอธิบายกฎความสัมพันธ์ของทั้ง 4 กลุ่ม ดังนี้

กลุ่มที่ 1 ด้านการบริการที่เป็นเลิศ (Excellent service หรือ Cluster E) สามารถอธิบายกฎความสัมพันธ์ที่ดีที่สุดได้ 4 กฎ ซึ่งสะท้อนให้เห็นรูปแบบความสัมพันธ์ของรายการคำถามของการท่องเที่ยวชุมชน ได้แก่ ความปลอดภัย (Security) อาหาร (Food) ร้านอาหาร (Restaurant) และสินค้าและบริการ (Products & Services) มี 4 กฎ สามารถอธิบายกฎความสัมพันธ์ได้ดังต่อไปนี้

กฎที่ 1: เมื่อนักท่องเที่ยวถามถึงมาตรฐานความสะอาด สินค้า-บริการที่หลากหลาย และสินค้า-บริการที่เพียงพอ ต้องถามถึงเจ้าหน้าที่ดูแลความปลอดภัยเสมอ มีค่าความเชื่อมั่น 99%

กฎที่ 2: เมื่อนักท่องเที่ยวถามถึงราคาสินค้า สินค้า-บริการที่หลากหลาย และสินค้า-บริการที่เพียงพอ ต้องถามถึงเจ้าหน้าที่ดูแลความปลอดภัยเสมอ มีค่าความเชื่อมั่น 98%

กฎที่ 3: เมื่อนักท่องเที่ยวถามถึงร้านอาหาร สินค้า-บริการที่หลากหลาย และสินค้า-บริการที่เพียงพอ ต้องถามถึงเจ้าหน้าที่ดูแลความปลอดภัยเสมอ มีค่าความเชื่อมั่น 98%

กฎที่ 4: เมื่อนักท่องเที่ยวถามถึงอาหารเฉพาะถิ่น สินค้า-บริการที่หลากหลาย และสินค้า-บริการที่เพียงพอ ต้องจะถามถึงเจ้าหน้าที่ดูแลความปลอดภัยเสมอ มีค่าความเชื่อ

กลุ่มที่ 2 ด้านสิ่งอำนวยความสะดวกที่ได้มาตรฐาน (Standard facilities หรือ Cluster S) สามารถอธิบายกฎความสัมพันธ์ที่ดีที่สุดได้ 4 กฎ ซึ่งสะท้อนให้เห็นรูปแบบความสัมพันธ์ของรายการคำถามของการท่องเที่ยวชุมชน ได้แก่ ความปลอดภัย (Security) และ กิจกรรมสำหรับการท่องเที่ยว (Activities) มี 4 กฎ สามารถอธิบายกฎความสัมพันธ์ได้ดังต่อไปนี้

กฎที่ 1: เมื่อนักท่องเที่ยวถามถึงกิจกรรมการท่องเที่ยวสามารถสะท้อนเอกลักษณ์วัฒนธรรม วิถีชีวิต และภูมิปัญญาท้องถิ่น ต้องถามถึงเจ้าหน้าที่ดูแลความปลอดภัยเสมอ มีค่าความเชื่อมั่น 100%

กฎที่ 2: เมื่อนักท่องเที่ยวถามถึงเจ้าหน้าที่ดูแลความปลอดภัย ต้องถามถึงกิจกรรมการท่องเที่ยวสามารถสะท้อนเอกลักษณ์วัฒนธรรม วิถีชีวิต และภูมิปัญญาท้องถิ่นเสมอ มีค่าความเชื่อมั่น 100%

กฎที่ 3: เมื่อนักท่องเที่ยวถามถึงกิจกรรมการท่องเที่ยวสร้างความรู้ความเข้าใจในเอกลักษณ์ของชุมชน ต้องถามถึงเจ้าหน้าที่ดูแลความปลอดภัยเสมอ มีค่าความเชื่อมั่น 100%

กฎที่ 4: เมื่อนักท่องเที่ยวถามถึงกิจกรรมการท่องเที่ยวสร้างความรู้ความเข้าใจในเอกลักษณ์ของชุมชน ต้องถามถึงกิจกรรมการท่องเที่ยวสามารถสะท้อนเอกลักษณ์วัฒนธรรม วิถีชีวิต และภูมิปัญญาท้องถิ่นเสมอ มีค่าความเชื่อมั่น 100%

กลุ่มที่ 3 ด้านข้อมูลสารสนเทศที่ถูกต้องและตรงกับความต้องการ (Accuracy of information หรือ Cluster A) สามารถอธิบายความสัมพันธ์ที่ดีที่สุดได้ 3 กฎ ซึ่งสะท้อนให้เห็นรูปแบบความสัมพันธ์ของรายการคำถามของการท่องเที่ยวชุมชน ได้แก่ การให้คำแนะนำและบริการ (Suggestions & Service) และ ช่องทางการสื่อสารประชาสัมพันธ์ (Public relations, channels to communicate, and social media) มี 4 กฎ สามารถอธิบายความสัมพันธ์ได้ดังต่อไปนี้

กฎที่ 1: เมื่อนักท่องเที่ยวถามถึงระเบียบ ข้อปฏิบัติ เว็บไซต์ การประชาสัมพันธ์ คู่มือ และเอกสารเผยแพร่ ต้องถามถึงเจ้าหน้าที่ให้คำแนะนำเสมอ มีค่าความเชื่อมั่น 99%

กฎที่ 2: เมื่อนักท่องเที่ยวถามถึงระเบียบ ข้อปฏิบัติ เว็บไซต์ สื่อโซเชียลมีเดีย คู่มือและเอกสารเผยแพร่ ต้องถามถึงเจ้าหน้าที่ให้คำแนะนำเสมอ มีค่าความเชื่อมั่น 99%

กฎที่ 3: เมื่อนักท่องเที่ยวถามถึงระเบียบและข้อปฏิบัติ เว็บไซต์ สื่อโซเชียลมีเดีย การประชาสัมพันธ์ ต้องถามถึงเจ้าหน้าที่ให้คำแนะนำเสมอ มีค่าความเชื่อมั่น 99%

กลุ่มที่ 4 ด้านวัฒนธรรมที่โดดเด่น (Noble culture หรือ Cluster N) สามารถอธิบายความสัมพันธ์ที่ดีที่สุดได้ 4 กฎ ซึ่งสะท้อนให้เห็นรูปแบบความสัมพันธ์ของรายการคำถามของการท่องเที่ยวชุมชน ได้แก่ ประวัติศาสตร์ (Historical) การเข้าถึงแหล่งท่องเที่ยว (Tourist attractions) อัตลักษณ์ของชุมชน (Identity) และการอนุรักษ์ไว้ซึ่งประเพณีท้องถิ่น มี 4 กฎ สามารถอธิบายความสัมพันธ์ได้ดังต่อไปนี้

กฎที่ 1: เมื่อนักท่องเที่ยวถามถึงแหล่งท่องเที่ยว อัตลักษณ์เฉพาะ เอกลักษณ์โดดเด่น วัฒนธรรม ต้องถามประวัติศาสตร์ วิถีชีวิต ประเพณีเสมอ มีค่าความเชื่อมั่น 100%

กฎที่ 2: เมื่อนักท่องเที่ยวถามถึงอัตลักษณ์เฉพาะ เอกลักษณ์โดดเด่น วัฒนธรรม ต้องถามถึงประวัติศาสตร์ วิถีชีวิต และประเพณีเสมอ มีค่าความเชื่อมั่น 99%

กฎที่ 3: เมื่อนักท่องเที่ยวถามถึงแหล่งท่องเที่ยว อัตลักษณ์เฉพาะ วัฒนธรรม ต้องถามถึงประวัติศาสตร์ วิถีชีวิต และประเพณีเสมอ มีค่าความเชื่อมั่น 99%

4.5 สรุปผลการทดลอง

สรุปผลการวิเคราะห์กฎความสัมพันธ์รายการคำถามที่ถูกถามบ่อย แสดงให้เห็นถึงพฤติกรรมการสอบถามของนักท่องเที่ยวในพื้นที่การท่องเที่ยวโดยชุมชนเขตอีสานใต้ เพื่อนำผลลัพธ์จากขั้นตอนนี้ไปเป็น ข้อสอบถามหรือข้อมูลนำเข้าในวัตถุประสงค์ข้อที่ 3 ซึ่งเป็นวัตถุประสงค์สุดท้ายของงานวิจัย สามารถสร้างกฎความสัมพันธ์และรูปแบบรายการคำถามที่มี

ความสัมพันธ์กับการท่องเที่ยวอีสานใต้ นำเสนอด้วยผังภาพเพื่อให้ง่ายต่อการทำความเข้าใจ โดยพิจารณาจากคำสำคัญ (Keywords) ของทั้ง 4 กลุ่ม ผลลัพธ์แสดงดังตารางที่ 4-7

ตารางที่ 4 รูปแบบรายการคำถามที่มีความสัมพันธ์กับการท่องเที่ยวอีสานใต้ Class E

Class E	Question pattern
Rule 1	แหล่งท่องเที่ยวชุมชนมีห้องน้ำสะอาดมาตรฐานเพียงพอร้านค้าสินค้าบริการหลากหลายและเจ้าหน้าที่ดูแลความปลอดภัยนักท่องเที่ยว
Rule 2	แหล่งท่องเที่ยวชุมชนมีร้านค้ากำหนดราคาสินค้าบริการเหมาะสมหลากหลายเพียงพอและเจ้าหน้าที่ดูแลความปลอดภัยนักท่องเที่ยว
Rule 3	แหล่งท่องเที่ยวชุมชนมีร้านอาหารสะอาดได้มาตรฐานเพียงพอร้านค้าสินค้าบริการหลากหลายและเจ้าหน้าที่ดูแลความปลอดภัยนักท่องเที่ยว
Rule 4	แหล่งท่องเที่ยวชุมชนมีอาหารเอกลักษณ์บริการร้านค้าสินค้าหลากหลายเพียงพอและเจ้าหน้าที่ดูแลความปลอดภัยนักท่องเที่ยว

ตารางที่ 5 รูปแบบรายการคำถามที่มีความสัมพันธ์กับการท่องเที่ยวอีสานใต้ Class S

Class S	Question pattern
Rule 1	แหล่งท่องเที่ยวชุมชนมีกิจกรรมสะท้อนเอกลักษณ์วัฒนธรรมวิถีชีวิตภูมิปัญญาท้องถิ่นและที่พักปลอดภัย
Rule 2	แหล่งท่องเที่ยวชุมชนมีโรงแรมที่พักปลอดภัยและกิจกรรมสะท้อนเอกลักษณ์วัฒนธรรมวิถีชีวิตภูมิปัญญาท้องถิ่น
Rule 3	แหล่งท่องเที่ยวชุมชนมีกิจกรรมท่องเที่ยวความรู้ความเข้าใจเอกลักษณ์ชุมชนและโรงแรมที่พักปลอดภัย
Rule 4	แหล่งท่องเที่ยวชุมชนมีกิจกรรมท่องเที่ยวความรู้ความเข้าใจเอกลักษณ์ชุมชนและกิจกรรมสะท้อนวัฒนธรรมวิถีชีวิตภูมิปัญญาท้องถิ่น

ตารางที่ 6 รูปแบบรายการคำถามที่มีความสัมพันธ์กับการท่องเที่ยวอีสานใต้ Class A

Class A	Question pattern
Rule 1	แหล่งท่องเที่ยวชุมชนแจ้งระเบียบข้อปฏิบัติมีระบบสารสนเทศเว็บไซต์ข้อมูลประชาสัมพันธ์กิจกรรมการท่องเที่ยวช่องทางสังคมออนไลน์มีคู่มือเอกสารเผยแพร่มีเจ้าหน้าที่แนะนำบริการมีคุณภาพกับนักท่องเที่ยว
Rule 2	แหล่งท่องเที่ยวชุมชนแจ้งระเบียบข้อปฏิบัติมีระบบสารสนเทศเว็บไซต์ข้อมูลช่องทางการสื่อสารแฟนเพจเฟซบุ๊กไลน์คู่มือเอกสารเผยแพร่มีเจ้าหน้าที่แนะนำบริการมีคุณภาพกับนักท่องเที่ยว
Rule 3	แหล่งท่องเที่ยวชุมชนแจ้งระเบียบข้อปฏิบัติมีระบบสารสนเทศเว็บไซต์ข้อมูลช่องทางการสื่อสารแฟนเพจเฟซบุ๊กไลน์ประชาสัมพันธ์กิจกรรมการท่องเที่ยวช่องทางสังคมออนไลน์มีเจ้าหน้าที่แนะนำบริการมีคุณภาพกับนักท่องเที่ยว

ตารางที่ 7 รูปแบบรายการคำถามที่มีความสัมพันธ์กับการท่องเที่ยวอีสานใต้ Class N

Class N	Question pattern
Rule 1	ชุมชนมีแหล่งท่องเที่ยวหลากหลายอัตลักษณ์เฉพาะท้องถิ่นเอกลักษณ์โดดเด่นอนุรักษ์วัฒนธรรมท้องถิ่น ประวัติศาสตร์วัฒนธรรมวิถีชีวิตและประเพณี
Rule 2	ชุมชนมีแหล่งท่องเที่ยวอัตลักษณ์เฉพาะท้องถิ่นเอกลักษณ์โดดเด่นอนุรักษ์วัฒนธรรมท้องถิ่น ประวัติศาสตร์วัฒนธรรม วิถีชีวิตและประเพณี
Rule 3	ชุมชนมีแหล่งท่องเที่ยวหลากหลายอัตลักษณ์เฉพาะท้องถิ่นอนุรักษ์วัฒนธรรมท้องถิ่น ประวัติศาสตร์วัฒนธรรม วิถีชีวิตและประเพณี

ด้านการบริการที่เป็นเลิศ พบว่าคุณภาพสัมพันธ์และรูปแบบพฤติกรรมกรรมการสอบถามที่ปรากฏร่วมกันบ่อยของนักท่องเที่ยวในประเด็นความปลอดภัย โดยเริ่มคำถามแรกคือ ความสะอาดและห้องน้ำที่ได้มาตรฐาน สินค้าและบริการที่หลากหลาย สินค้าและบริการที่เพียงพอ ราคาของสินค้า ร้านอาหารที่ถูกสุขอนามัย อาหารพื้นถิ่น

ด้านสิ่งอำนวยความสะดวกที่มีมาตรฐาน พบว่าคุณภาพสัมพันธ์และรูปแบบพฤติกรรมกรรมการสอบถามที่ปรากฏร่วมกันบ่อยของนักท่องเที่ยวยังให้ความสนใจในประเด็นความปลอดภัย โดยเริ่มตั้งคำถามจาก กิจกรรมการท่องเที่ยวสามารถสะท้อนเอกลักษณ์วัฒนธรรม วิถีชีวิต และภูมิปัญญาท้องถิ่น กิจกรรมการท่องเที่ยวสร้างความรู้ความเข้าใจในเอกลักษณ์ของชุมชน

ข้อมูลสารสนเทศที่ถูกต้องและตรงกับความต้องการ พบว่าคุณภาพสัมพันธ์ที่นักท่องเที่ยวให้ความสนใจถามคือ เจ้าหน้าที่ให้คำแนะนำและการบริการที่ดีและมีคุณภาพ เริ่มต้นคำถามจากระเบียบ กฎและกติกา ช่องทางการสื่อสารเช่น เว็บไซต์ สื่อสังคมออนไลน์ คู่มือการท่องเที่ยว

อัตลักษณ์ ประเพณีและวัฒนธรรม การสอบถามที่พบบ่อยของนักท่องเที่ยวจะให้ความสำคัญในประเด็น แหล่งท่องเที่ยวทางประวัติศาสตร์ที่สะท้อนวัฒนธรรม วิถีชีวิต และประเพณีของชุมชน โดยคำถามของนักท่องเที่ยวจะเริ่มจาก การเข้าถึงแหล่งท่องเที่ยว อัตลักษณ์ เอกลักษณ์ การส่งเสริมและอนุรักษ์วัฒนธรรมท้องถิ่น

บทที่ 5

การพัฒนากรอบการทำงานการหาคำตอบสำหรับตอบคำถามอัตโนมัติการท่องเที่ยว โดยชุมชนอีสานใต้

การดำเนินการตามวัตถุประสงค์ที่สาม เพื่อพัฒนากรอบการทำงานการหาคำตอบสำหรับตอบคำถามอัตโนมัติการท่องเที่ยวโดยชุมชนอีสานใต้ ด้วยแนวคิดการค้นคืนเอกสาร (Information retrieval) โดยผู้วิจัยได้ศึกษาเอกสารและผลงานวิจัยที่เกี่ยวข้องกับสาระสำคัญ เพื่อให้เกิดความเข้าใจที่ชัดเจน ตรงประเด็นได้มาซึ่งประโยชน์ในเชิงวิชาการที่แท้จริง โดยนำเสนอใน 5 ประเด็นสำคัญ ได้แก่ 1) ที่มาและความสำคัญของปัญหา 2) แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 3) ระเบียบวิธีวิจัย 4) ผลการทดลอง และ 5) สรุปผลการทดลอง รายละเอียดตามลำดับดังนี้

5.1 ที่มาและความสำคัญของปัญหา

การพัฒนาการท่องเที่ยวด้วยความยั่งยืน [4] เป็นหนึ่งในเป้าหมายขององค์การสหประชาชาติ (United nations: UN) และแผนแม่บทของโลก AGENDA 21 [1] [2] [3] จากสถิตินักท่องเที่ยวต่างชาติทั่วโลก ในปี 2560 ประเทศไทยสร้างรายได้มากที่สุดในภูมิภาคคิดเป็นเงิน 57.5 พันล้านดอลลาร์สหรัฐ อุตสาหกรรมท่องเที่ยวนับได้ว่าเป็นอุตสาหกรรมสำคัญที่สร้างงานและรายได้ให้กับคนไทย โดยแผนพัฒนาเศรษฐกิจและสังคมแห่งชาติ [5] ให้ความสำคัญกับการท่องเที่ยวเน้นคุณค่าและความยั่งยืน สอดคล้องกับยุทธศาสตร์หลักในการกระจายรายได้สู่การท่องเที่ยวเมืองรอง เพื่อกระตุ้นให้เกิดการท่องเที่ยวและสร้างเศรษฐกิจฐานรากอย่างต่อเนื่อง Community-based tourism: CBT [7] เป็นเครื่องมือสร้างความเข้มแข็งให้ชุมชนและสิ่งแวดล้อมอย่างยั่งยืน [8] โดยพบว่าแนวโน้มการท่องเที่ยวโดยชุมชนให้ความสำคัญกับนวัตกรรมการสื่อสารและเทคโนโลยี เพื่อเป็นเครื่องมือช่วยเหลือหรือให้ประสบการณ์เฉพาะด้านการท่องเที่ยว การเข้าถึงข้อมูลอย่างรวดเร็วและตรงกับความต้องการมากที่สุด อาทิเช่น ระบบแนะนำ (Recommendation system: RS) หรือ ระบบถามตอบอัตโนมัติ (Question answering system: QAs) อย่างไรก็ตามปัญหาสำคัญของอุตสาหกรรมการท่องเที่ยวยังไม่ได้รับการแก้ไขให้ตรงจุด อาทิเช่น 1) ปัญหาความความเลื่อมล้ำในสังคม [10] การพัฒนาอย่างก้าวกระโดดของเทคโนโลยี 2) การประชาสัมพันธ์ไม่ตรงกับกลุ่มเป้าหมาย [12] โดยเน้นแหล่งท่องเที่ยวหลักมากกว่าแหล่งท่องเที่ยวรองและขาดการส่งเสริมในเชิงพาณิชย์อย่างจริงจัง และ 3) การบริหารจัดการการเข้าถึงข้อมูลข่าวสารในเชิงพื้นที่ การเตรียมสารสนเทศที่รวดเร็วและตรงกับความต้องการ ซึ่งปัจจุบันเทคโนโลยีที่ถูกพัฒนาขึ้น [15] ยังไม่สามารถตอบโจทย์ได้ มากไปกว่านั้นคือข้อมูลจำนวนมหาศาลที่ได้จากการสืบค้นข้อมูลด้วย Search

engine คือข้อจำกัดในการเข้าถึงข้อมูล อาทิเช่น 1) การดึงและแสดงผลข้อมูลที่ไม่เกี่ยวข้อง 2) ข้อมูลมีจำนวนมากยากต่อการนำไปใช้ประโยชน์ 3) ยากต่อการตรวจสอบของแหล่งที่มาและข้อเท็จจริง และ 4) ขาดความแม่นยำและไม่ตรงกับความต้องการ ดังนั้นการพัฒนาเครื่องมือค้นหาจึงจำเป็นต้องมีคุณสมบัติพิเศษในการดึงเอกสารจากลักษณะเด่นของข้อมูล [69] ด้านการทอ่งเทียวกได้ อย่างมีประสิทธิภาพ วัตถุประสงค์ของงานวิจัยในครั้งนี้เพื่อพัฒนารอบการทำงานการหาความสัมพันธ์คำถามคำตอบอัตโนมัติ ในการสร้างต้นแบบและแก้ปัญหาการเข้าถึงข้อมูลของนักทอ่งเทียวในเชิงพื้นที่กรณีสึกษาอิสานใต้ ด้วยแนวคิดการค้นคืนข้อมูล (Information retrieval model) ได้ผลลัพธ์ที่รวดเร็ว ถูกต้องและแม่นยำ สร้างความมั่นใจและโอกาสต่อการตัดสินใจเดินทางของนักทอ่งเทียว [7] [11] [9] [21] กระตุ้นให้เกิดการทอ่งเทียวและสร้างเศรษฐกิจฐานราก

5.2 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยในครั้งนี้นำเสนอแนวคิด ทฤษฎีและงานวิจัยที่เป็นประเด็นสำคัญ โดยแบ่งออกเป็น 2 ส่วนหลัก คือ ส่วนแรก แนวคิดและทฤษฎีที่เกี่ยวข้อง ประกอบด้วย การค้นคืนสารสนเทศ (Information retrieval) แบบจำลองปริภูมิเวกเตอร์ (Vector space model: VSM) การตัดคำภาษาไทย (Thai words segmentation) การให้ค่าน้ำหนัก (Term weighting) การวัดความคล้ายคลึง (Similarity measurement) และการประเมินประสิทธิภาพการค้นคืนสารสนเทศ (Information retrieval model evaluation) และส่วนที่สอง คือ งานวิจัยที่เกี่ยวข้อง ดังรายละเอียดต่อไปนี้

5.2.1 แนวคิดและทฤษฎีที่เกี่ยวข้อง

1) การค้นคืนสารสนเทศ (Information retrieval: IR)

การค้นคืนสารสนเทศ [17] [70] เป็นศาสตร์แขนงหนึ่งของ Natural language processing (NLP) และ Artificial intelligence (AI) ที่ช่วยให้คอมพิวเตอร์เข้าใจภาษาธรรมชาติของมนุษย์ สิ่งแรกที่จะต้องจัดการคือทำให้คอมพิวเตอร์รู้จักคำในภาษาไทย โดยการเสนอข้อมูล (Representation) การจัดการข้อมูล (Organization) และการเข้าถึงข้อมูล (Access) จำนวนมหาศาลซึ่งถูกเก็บไว้ในรูปแบบดิจิทัล [70] ระบบค้นคืนสารสนเทศ แบ่งได้ 2 ประเภท คือ 1) การเรียกดูสารสนเทศ (Information browsing) และ 2) การค้นคืนสารสนเทศ (Information Retrieval) โดยแตกแขนงได้อีก 2 ประเภท คือ ประเภทที่หนึ่ง คือ Classic information retrieval models ทำงานบนพื้นฐานของตัวดำเนินการตรรกศาสตร์ (Boolean) เวกเตอร์ (Vector) หรือความน่าจะเป็น (Probabilistic) ประเภทที่สอง คือแบบ Structural model จะทำงานบนพื้นฐานโครงสร้างของข้อมูล เช่น Tree and ontology

[42] งานวิจัยในครั้งนี้นำแนวคิดการค้นคืนข้อความ (Textual information retrieval) ข้อมูลนำเข้าเป็นเอกสารที่ไม่มีโครงสร้าง ดำเนินการวิเคราะห์ บันทึกจัดเก็บและดึงเอกสารจากลักษณะเด่นของข้อมูลแบบอัตโนมัติ [69] ด้วยการดึงเอกสารที่เกี่ยวข้อง (Relevant) กับข้อความคำถามและแยกเอกสารที่ไม่เกี่ยวข้อง (Non-relevant) เพื่อให้ผลการค้นคืนตรงกับความต้องการของผู้ใช้บนพื้นฐานแบบจำลอง VSM หรือ Term vector model

2) แบบจำลองปริภูมิเวกเตอร์ (Vector space model: VSM)

แบบจำลองปริภูมิเวกเตอร์ ใช้สำหรับอธิบายโมเดลของเอกสารแบบ Term-document matrix (TDM) และสร้างดัชนีสำหรับการค้นคืน ซึ่งแต่ละหน่วยใน Matrix คือค่าน้ำหนักที่ขึ้นอยู่กับ Term frequency ที่พบในเอกสารแล้วเปรียบเทียบกับเอกสารทั้ง 2 ฉบับ ผลลัพธ์ของคำสอบถามกับเอกสารคำตอบจะถูกจัดลำดับจากค่าและความเหมือน ดังนั้น TDM จึงเป็นพื้นฐานสำคัญของการค้นคืนเอกสาร โดยกล่าวถึงจุดเด่นและจุดด้อยของโมเดลพื้นฐาน ทั้ง 4 ประเภท ได้แก่ 1) Boolean model 2) Vector space model 3) Probabilistic model and 4) Hybrid model แสดงดังตารางที่ 8

ตารางที่ 8 วิเคราะห์จุดเด่น-ด้อย วิธีการค้นคืนสารสนเทศ

ประเภทโมเดลพื้นฐาน	วิธีการ	จุดเด่น	จุดด้อย
Boolean model [71][72][73]	ทำงานบนแนวคิดเชิงตรรกะ	ง่าย นิยมใช้ และ มีประสิทธิภาพ ใช้กับข้อมูลไม่มาก	ไม่จัดลำดับ ไม่เปรียบเทียบ
Vector space model [18]	ให้ความสำคัญกับความถี่ของคำ และการให้น้ำหนักคำ	ง่าย ใช้กับข้อมูลมาก จัดลำดับ	ไม่สนใจความหมายคำ สืบค้นแบบเงื่อนไขไม่ได้
Probabilistic model	ใช้ความน่าจะเป็นโดยคำนวณ จากข้อมูลความถี่ของคำ	ง่าย ใช้กับข้อมูลมาก จัดลำดับ	ไม่สนใจความหมายคำ สืบค้นแบบเงื่อนไขไม่ได้
Hybrid model [19][74][75]	ปรับปรุงข้อด้อยของ Boolean และ Vector space model	ให้น้ำหนักของคำ เชื่อมต่อทางตรรกะได้	-

2.1) Boolean algebra ใช้แนวความคิดเชิงตรรกะ หรือพีชคณิตบูลีน (Boolean algebra) เช่น AND (ทั้งสองคำที่ถูกเชื่อม) OR (อย่างใดอย่างหนึ่ง) และ NOT เมื่อพัฒนาระบบค้นคืน โมเดลของเอกสารจะมี Term-document และสร้างดัชนีเพื่ออำนวยความสะดวกในการค้นหา เมื่อผู้ใช้งานกำหนดคำสอบถามจะเข้าสู่การ Match เทอมในคำสอบถาม ผลลัพธ์ที่ได้คือ ใช่ และไม่ใช่ ต่อมาได้มีการพัฒนาโมเดลเพิ่มเติมคือ 1) Extended boolean model และ 2) Fuzzy set model

2.2) Vector space model หรือ Matrix algebra โมเดลของเอกสารจะเป็นรูป Term-document matrix และสร้างดัชนีสำหรับการค้นคืน โดยแต่ละหน่วยใน Matrix คือค่าน้ำหนักที่ขึ้นอยู่กับความถี่ของเทอม (Term frequency) ที่พบในเอกสาร แล้วเปรียบเทียบกับเอกสาร 2 ฉบับ ให้

ความสำคัญกับความถี่และน้ำหนักของคำ ผลลัพธ์ของคำสอบถามกับเอกสารคือ หากทำมุมกันใกล้ 0° ค่า $\cos \theta$ จะเข้าใกล้ 1 (พบเอกสาร) และ หากทำมุมใกล้ 90° ค่า $\cos \theta$ จะเข้าใกล้ 0 (ไม่พบเอกสาร) แล้วจัดลำดับเอกสารจากค่าและความเหมือน ต่อมาได้มีการพัฒนาโมเดลเพิ่มเติมคือ 1) Generalized vector space model 2) Topic-base vector space model 3) Latent semantic indexing model และ 4) Neural network model

2.3) Probabilistic model ใช้แนวคิดความน่าจะเป็น (Probability theory) สามารถอธิบายได้ว่าในโมเดลจะ Matching คำสอบถามกับเอกสารเพื่อเปรียบเทียบความน่าจะเป็น โดยคำนวณจากข้อมูลความถี่ของคำ พิจารณาใน 2 ค่า คือ

$P(R | q, d_j)$ คือความน่าจะเป็นที่เอกสารจะเกี่ยวข้องกับคำสอบถาม

$P(\bar{R} | q, d_j)$ คือความน่าจะเป็นที่เอกสารจะไม่เกี่ยวข้องกับคำสอบถาม

ดังนั้น $P(R | q, d_j) / P(\bar{R} | q, d_j)$ เป็นสมการที่กำหนด แต่ยังไม่พบข้อจำกัดที่ความซับซ้อนทำให้ไม่ได้รับความนิยม ต่อมาได้มีการพัฒนาโมเดลเพิ่มเติมคือ 1) Inference network model 2) Belief network model

3) การตัดคำภาษาไทย (Thai words segmentation)

การตัดคำภาษาไทย (Thai words segmentation) [76] เป็นกระบวนการพื้นฐานของ NLP และความท้าทายของการประมวลผลภาษาไทย คือการทำให้คอมพิวเตอร์สามารถเข้าใจภาษาธรรมชาติของมนุษย์ได้อย่างมีประสิทธิภาพ ปัญหาและอุปสรรคที่พบคือการเขียนติดกันทั้งประโยคโดยไม่มีเครื่องหมายวรรคตอนหรือแสดงการแบ่งคำดังเช่นภาษาอังกฤษ การตัดคำหรือการแบ่งคำที่ถูกต้องจึงมีความสำคัญกับภาษาไทย เพื่อนำข้อมูลคำศัพท์ไปใช้ประโยชน์ต่อ ซึ่งส่งผลกระทบต่อประมวลผลของคอมพิวเตอร์และการค้นหาคำทำได้ถูกต้องและแม่นยำ ปัจจุบันมีหลายเทคนิคที่ได้รับความนิยมและสามารถรองรับการตัดคำภาษาไทย เช่น 1) Longest matching 2) Probabilistic model 3) Feature-based approach และ 4) Maximal matching (newmm) [77] [44] [45] อัลกอริทึมนี้ถูกนำมาใช้แก้ปัญหาการตัดคำภาษาไทยใน PyThaiNLP tokenize ได้อย่างรวดเร็วและถูกต้อง ด้วยหลักการ Dictionary-based (DCB) ฐานข้อมูลคำศัพท์ (WordNet) [78] ที่รวมกลุ่มคำศัพท์ไวยากรณ์ (Synset) หรือเซตของคำที่มีความหมายเหมือนกันไว้ โดยสามารถสรุปกระบวนการของ newmm [43] ดังตัวอย่างข้อความ “ไปหามเหสี” สามารถแบ่งคำได้ 2 แบบ คือ แบบที่หนึ่ง “ไป-หาม-เห-สี” แบ่งคำได้ 4 คำ และ แบบที่สอง “ไป-หา-มเหสี” แบ่งคำได้ 3 คำ อธิบายใน 2 ขั้นตอนดังนี้ ขั้นที่ 1 เลือกผลลัพธ์การตัดคำมีจำนวนค่าน้อยที่สุด ผลลัพธ์ที่ได้คือ แบบที่สอง “ไป-หา-มเหสี” และ ขั้นที่ 2 หากมีจำนวนคำเท่ากันจะย้อนกลับ (Backtracking) ไปใช้วิธีการ Longest matching

4) การให้ค่าน้ำหนัก (Term weighting)

การให้ค่าน้ำหนัก (Term weighting) เป็นวิธีการกำหนดค่าน้ำหนักสำหรับคำที่มีความสำคัญ (keywords) หรือ document representation เพื่อบอกความสำคัญของคำ (Term) ที่ใช้สำหรับการบ่งชี้ (Index) เอกสารฉบับนั้น โดยการให้น้ำหนักคำ ๆ หนึ่งในเอกสารฉบับหนึ่งจะพิจารณาจากความถี่ของคำ (Term frequency) ที่ปรากฏในเอกสารนั้นและจำนวนของเอกสารทั้งหมดที่มีคำนั้นปรากฏอยู่ วิธีการให้น้ำหนักของคำวิธีหนึ่งที่ได้รับการนิยามคือ TF-IDF

4.1) TF-IDF (Term frequency - inverse dense frequency) [79] [80] [81] เป็นเทคนิคพื้นฐานที่ใช้ในการวิเคราะห์ข้อความทำงานบนพื้นฐานของเมตริก โดยการสกัดเอาข้อมูลที่สำคัญออกมาจากข้อความในเอกสาร ด้วยหลักการพิจารณาองค์ประกอบของคำภายในประโยคของเอกสาร หรือการดึงเอกสารจากลักษณะเด่นจากเนื้อหาเอกสารเพื่อเป็นตัวแทนของเอกสาร ขณะเดียวกันเทคนิคนี้จะไม่สามารถอธิบายความหมายและความแตกต่างของเอกสารที่เกิดขึ้นร่วมกันได้ โดยจะไม่นำลำดับของคำในเอกสารมาพิจารณา มีองค์ประกอบหลัก 2 ส่วน ได้แก่

4.1.1) Term frequency (TF) เป็นค่าที่บอกความถี่ของคำแต่ละคำที่พบในเอกสาร โดยคำนวณจากการนำจำนวนครั้งที่คำนั้นๆ ปรากฏในเอกสารมาหารด้วยจำนวนคำทั้งหมดในเอกสาร แสดงดังสมการที่ 5.1

Term frequency (TF) หรือความถี่ของคำศัพท์ที่ปรากฏในเอกสารฉบับนั้น

$$TF(t, d) = \frac{\text{จำนวนคำที่ปรากฏในเอกสาร}}{\text{จำนวนคำทั้งหมดในเอกสาร}} \quad (5.1)$$

เมื่อ TF(t, d) is the number of occurrences of “t” in document “d”

t คือ term หรือ คำ

d คือ เอกสารฉบับนั้น

4.1.2) Inverse document frequency (IDF) เป็นการคำนวณค่าน้ำหนัก (Weight) ความสำคัญของแต่ละคำ ซึ่งคำที่พบบ่อยในทุกเอกสาร จะมีค่า IDF ต่ำ ส่งผลให้ค่าเหล่านั้นไม่มีลักษณะเด่นและเป็นตัวแทนของเอกสารไม่ได้ ค่า IDF สามารถคำนวณได้จากสมการที่ 5.2

Inverse document frequency (IDF) หรือ การผกผันในความถี่ของเอกสาร

$$IDF(t) = \log \left[\frac{(1+n)}{(1+DF(t))} \right] + 1 \quad (5.2)$$

เมื่อ t คือ term หรือคำ 1 คำ

n คือ จำนวนเอกสารทั้งหมดที่มีอยู่

DF(t) คือ จำนวนเอกสารที่พบของคำ t

ดังนั้น **TF-IDF** จะทำการหาค่าน้ำหนักของความถี่ของคำที่ปรากฏในเอกสารนั้น และค่าน้ำหนักการพกผันในความถี่ของเอกสารที่คำนวณจากเอกสารทั้งหมด โดยคำที่มีค่า TF-IDF สูงจึงมีแนวโน้มที่จะเป็นใจความสำคัญของเอกสาร ค่า TF-IDF สามารถคำนวณได้จากสมการที่ 3

$$TFIDF(t, d) = TF(t, d) * IDF(t) \quad (5.3)$$

เมื่อ t คือ term หรือ คำ
d คือ เอกสารฉบับนั้น

5) การวัดความคล้ายคลึง (Similarity measurement)

ความคล้ายคลึง [82] เป็นตัวบ่งชี้ความเหมือนด้วยวิธีการเปรียบเทียบจากการวัดระยะห่างระหว่างเอกสาร 2 เอกสารแล้วคำนวณออกมาเป็นตัวเลข โดยผลลัพธ์จะมีค่าอยู่ระหว่าง 0 ถึง 1 โดย 0 หมายถึงมีความคล้ายคลึงกันน้อยที่สุด และ 1 หมายถึงมีความคล้ายคลึงกันมากที่สุด ปัจจุบันการคำนวณหาค่าความคล้ายคลึงจากการวัดระยะห่างมีหลายวิธี [83] อาทิเช่น 1) Euclidean 2) Manhattan 3) Hamming 4) Jaccard coefficient และ 5) Cosine similarity ซึ่งเป็นวิธีที่ได้รับความนิยมและถูกนำมาใช้อย่างแพร่หลายสำหรับงานค้นคืนเอกสาร โดยพิจารณาจากทิศทางของเอกสารบนพื้นที่เวกเตอร์และการให้ค่าน้ำหนักของคำ (Term weight) มาวัดความคล้ายคลึงตามแนวคิดปริภูมิเวกเตอร์ (VSM) เพื่อหาค่าความเกี่ยวข้องในโครงสร้างของข้อความ (Syntactic text similarity) โดยให้ความสำคัญกับโครงสร้างและลำดับของคำแต่ไม่สนใจความหมาย

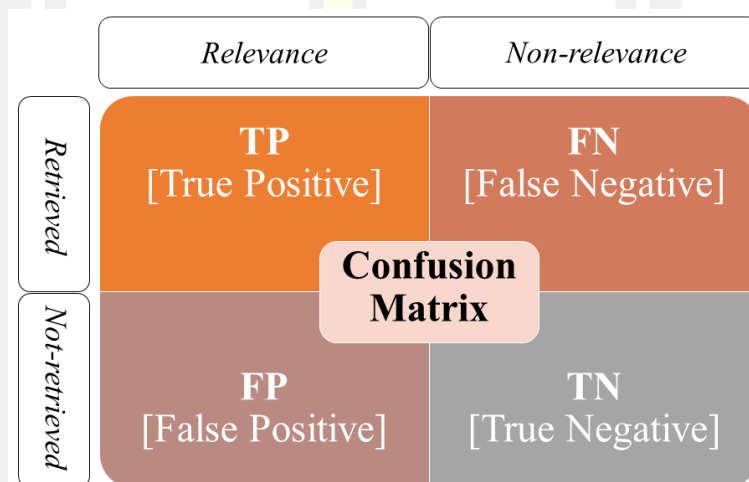
วัดค่าความคล้ายคลึงด้วยองศา (Cosine similarity) [84] [83] เป็นวิธีการทางสถิติเพื่อใช้วัดค่าความคล้ายคลึงด้วยองศาหรือระยะห่างของเวกเตอร์ระหว่างเอกสารทั้งสอง เป็นตัวชี้วัดค่าความคล้ายคลึงเชิงมุมโคไซน์ โดยที่ค่าโคไซน์ระหว่างมุมของเวกเตอร์ จะมีค่าอยู่ที่ 0 ถึง 1 ถ้าค่า Cosine มีค่าเข้าใกล้เดียวกับ 1 แสดงว่าเอกสารมีความคล้ายคลึงกันมาก และหากค่าเข้าใกล้เดียวกับ 0 จะมีความคล้ายคลึงกันน้อย ใช้วิธีการนำเวกเตอร์ของข้อมูลทั้งสองในปริภูมิ n มิติ มาทำการ Dot product เพื่อหาค่ามุมที่สองเวกเตอร์นั้น ๆ โดยถ้ามุมมีค่าน้อย คือสองเวกเตอร์มีความเหมือนกัน หากมุมมีค่ามาก หมายถึงสองเวกเตอร์มีความเหมือนกันน้อย โดยคำนวณได้จากสมการที่ 5.4

$$\text{cossim}(d_1, d_2) = \frac{d_1 \cdot d_2}{\|d_1\| \|d_2\|} = \frac{\sum_i (d_{i1} * d_{i2})}{\sqrt{d_{i1}^2} * \sqrt{d_{i2}^2}} \quad (5.4)$$

เมื่อ d_1 คือ เวกเตอร์ของเอกสารต้นฉบับ (Query)
 d_2 คือ เวกเตอร์ของเอกสารคำตอบ (Answer)
 i คือ จำนวนมิติคำในเวกเตอร์

6) การประเมินประสิทธิภาพการค้นคืนสารสนเทศ

การค้นคืนสารสนเทศ (Information retrieval) [16] [17] ที่ดีจะต้องมีลักษณะเด่นคือสามารถดึงเอาสารสนเทศที่มีความสัมพันธ์กับข้อสอบถามได้อย่างแม่นยำและจัดเรียงลำดับตามความสำคัญเพื่อสร้างความพึงพอใจต่อผู้ใช้งาน การประเมินความสามารถของแบบจำลองจึงเป็นสิ่งสำคัญในการอธิบายประสิทธิภาพของโมเดลที่สร้างขึ้น การประเมิน โมเดลในครั้งนี้ผู้วิจัยเลือกใช้ค่า Accuracy เพื่อประเมินความถูกต้องของการค้นคืน ซึ่งเป็นขั้นตอนสุดท้ายก่อนนำผลลัพธ์ที่ได้ไปใช้งานจริง โดยเปรียบเทียบผลลัพธ์ที่ได้จากโปรแกรมและผู้เชี่ยวชาญด้วยรูปแบบของเมตริกซ์วัดประสิทธิภาพ (Confusion Matrix) แสดงดังรูปที่ 20 และสมการที่ 5.5-5.8



รูปที่ 20 เมตริกซ์วัดประสิทธิภาพ (Confusion Matrix)

TP (True positive) หมายถึง สิ่งที่โปรแกรมทำนายว่าจริงและคนบอกว่าจริง

TN (True negative) หมายถึง สิ่งที่โปรแกรมทำนายว่าจริงและคนบอกว่าไม่จริง

FP (False positive) หมายถึง สิ่งที่โปรแกรมทำนายว่าไม่จริงและคนบอกว่าจริง และ

FN (False negative) หมายถึง สิ่งที่โปรแกรมทำนายว่าไม่จริงและคนบอกไม่ว่าจริง

$$Precision = \frac{TP}{TP+FP} \quad (5.5)$$

$$Recall = \frac{TP}{TP+FN} \quad (5.6)$$

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (5.7)$$

$$F_measure = \frac{2*Precision*Recall}{(Precision+Recall)} \quad (5.8)$$

5.2.2 งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้รวบรวมงานวิจัยที่เกี่ยวข้องบนพื้นฐานการสร้างตัวแทนเอกสาร (Document representation) ด้วยวิธีการคำนวณค่าน้ำหนัก (Term weighting) รวมถึงการประเมินความสามารถของแบบจำลองเป็นสิ่งสำคัญในการอธิบายประสิทธิภาพของโมเดลที่สร้างขึ้น ซึ่งเป็นที่นิยมและถูกนำมาประยุกต์ใช้งานในหลายมิติ อาทิเช่น อุตสาหกรรมท่องเที่ยว การตรวจจับข่าวปลอม และงานด้านอื่น ๆ รายละเอียดแสดงในตารางที่ 9

ตารางที่ 9 เปรียบเทียบวิธีการค้นคืนสารสนเทศและการประเมินประสิทธิภาพ

Aspects	Data collection	Similarity measure	Model evaluation				
			Method	Acc.	P	R	F1
hotel reviews [85]	515K.	TF-IDF, Cosine	-	-	-	-	
Hotel & restaurant reviews [86]	2000 reviews	TF-IDF	NB	92.5%	0.93	0.92	0.92
			SVM	81.8%	0.70	0.81	0.75
			RF	93.6%	0.93	0.93	0.93
online travel review [87]	20000 reviews	TF-IDF, Cosine	SVM	90.1%	0.85	0.90	0.87
Thai hotel reviews [88]	4129 reviews	TF-IDF, Cosine	Cosine	81.2%	-	-	-
Turkish news [89]	3600 news	TF-IDF, Cosine	k-NN	58%	0.58	0.58	0.58
Thai Fake News [90]	41,448 news	TF-IDF, Cosine	NB	78%	0.85	0.78	0.79
			SVM	96%	0.96	0.96	0.96
			RF	96%	0.97	0.96	0.96
News dataset [91]	1.02 GB	TF-IDF, Cosine	-	0.61	0.61	0.61	
Patents dataset [92]	102 pair	TF-IDF, Cosine	TFIDF	88%	-	-	-
			SVM	90%	-	-	-
Thai Subjective Exam. [93]	70 students enrolled	TF-IDF, Cosine	TFIDF	-	-	-	-
e-Book contents [94]	-	Jaccard, Cosine	DTD	-	0.75	1	0.85

อุตสาหกรรมท่องเที่ยว (Tourism industry) พบว่า Mishra และคณะ [85] ได้ประยุกต์ใช้คำศัพท์ (Term) จากถุงคำ (Bag of words) ด้วยวิธี TF-IDF และ Cosine similarity โดยรวบรวมข้อมูลการรีวิวโรงแรมในยุโรป จาก Blogs และ Social media เพื่อการเปรียบเทียบข้อมูลการจองและราคา ผลลัพธ์ที่ได้จากการทดลองพบว่าวิธีการดังกล่าวช่วยเพิ่มการตัดสินใจเลือกโรงแรมและที่พักสำหรับนักท่องเที่ยวได้อย่างมีประสิทธิภาพ และในปีเดียวกัน Afrizal และคณะ [86] ได้นำเทคนิค TF-IDF เพื่อใช้คัดกรองคุณลักษณะของข้อมูลความคิดเห็นจากการรีวิวโรงแรมและร้านอาหารของนักท่องเที่ยวจำนวน 2000 รีวิว ผลการทดลองพบว่า TF-IDF เป็นวิธีที่ดีที่สุดสำหรับ

คัดกรองคุณลักษณะของข้อมูลเมื่อเปรียบเทียบจากความถี่ของคำ (Term frequency) ของวิธีแบบ mTF-IDF และ BM25 โดยต่อมา Chen และคณะ [87] ใช้จุดเด่นของ TF-IDF วิเคราะห์ข้อมูลบนพื้นฐานของเมตริก จากความคิดเห็นใน Online travel บนเว็บไซต์ TripAdvisor จำนวน 20,000 Reviews และคำนวณค่า Cosine similarity ผลลัพธ์ที่ได้นักท่องเที่ยวสามารถนำไปใช้ประโยชน์สำหรับวางแผนการเดินทางได้ด้วยตนเอง เช่นเดียวกับ Khamphakdee และคณะ [88] ได้นำเสนอผลการวิจัยในปี 2021 วิเคราะห์ข้อมูลความคิดเห็นจาก Thai hotel reviews รวม 4000 รายการ โดยแบ่งข้อมูล 3129 รายการเป็น Training set และ 1000 รายการเป็น Test set ผลการวิจัยพบว่าวิธีการดังกล่าวสามารถสร้าง Text represent และหาค่าความคล้ายคลึงกันระหว่างข้อคิดเห็นจาก Corpus ที่ถูกสร้างขึ้นเองอย่างมีนัยสำคัญ ร้อยละ 81.2

การตรวจจับข่าวปลอม (Fake news) พบว่า Eminagaoglu และคณะ [89] นำเสนอแนวทางการตรวจจับข่าวปลอมระหว่างภาษาตุรกีและอังกฤษ จากชุดข้อมูลที่หลากหลายใน TTC-3600 Collections บนหลักการ TF-IDF นักวิจัยพบว่าการใช้วิธีการ Stemming algorithm มีผลต่อการได้มาซึ่งจำนวนของ Term ทำให้ส่งผลต่อความเกี่ยวข้อง (relevant) ของเอกสารและการตรวจจับข่าวปลอมอย่างมีนัยสำคัญ ต่อมา Meesad [90] ได้วิเคราะห์ข้อมูล Thai fake news จาก Web crawler เพื่อหาความถี่ของคำ จำนวนค่าน้ำหนัก (Weight) และความคล้ายคลึงกันระหว่างข้อความข่าว โดยแบ่งชุดข้อมูลสำหรับ Training 20,723 Validation 8,290 และ Testing 12,435 รวมทั้งสิ้น 41,448 นักวิจัยพบว่าเทคนิค TF-IDF และวิธีการ Cosine similarity มีผลต่อการตรวจจับข่าวปลอม ผลจากการทดลองพบว่าคำสอบถามและเนื้อหาของข่าวปลอมมีความคล้ายคลึงกัน สอดคล้องกับ Yang [91] ใช้จุดเด่นของ Text vectorize แล้วนำตัวเลขที่ได้มาคำนวณให้ได้คุณลักษณะสำคัญเพื่อเป็นตัวแทนเนื้อหาข่าวปลอมโดยไม่สนใจความหมายของคำ (Term) เพื่อเปรียบเทียบประสิทธิภาพการสกัดคุณลักษณะใน 3 โมเดล LSA LDA และ TF-IDF ผลการทดลองพบว่าวิธี TF-IDF สามารถสร้างตัวแทนของข่าวปลอมได้อย่างมีประสิทธิภาพมากที่สุด.

งานด้านอื่นที่นิยมนำเทคนิคค้นคืนไปประยุกต์ใช้งาน อาทิเช่น Shahmirzadi และคณะ [92] นำเสนอแนวคิดความคล้ายคลึงของข้อความ (Text similarity) บนหลักการ Vector space Model ด้วยเทคนิค TF-IDF เพื่อเปรียบเทียบความคล้ายคลึงของสิทธิบัตร (Patent-to-patent similarity) จำนวน 491 สิทธิบัตร (Patents) จากคลังข้อมูล USPTO ระหว่างปี 1976 - 2018 ผลการทดลองด้วยวิธี TF-IDF กับสิทธิบัตร (Patent) ให้ค่าความคล้ายคลึงสูงและให้ค่าประสิทธิภาพโดยรวม 98% ในปีต่อมา Saipetch และคณะ [93] ประยุกต์ใช้หลักการ TF-IDF เป็นตัวแทนเอกสารหาความเหมือนด้วยเทคนิคโคไซน์ (Cosine similarity) และการใช้คำไวพจน์ (Synonyms) จากการทำข้อสอบของนักศึกษา จำนวน 70 คน โดยการเปรียบเทียบคะแนนคำตอบของนักเรียนและ

ข้อคิดเห็นของผู้เชี่ยวชาญเป็นไปในทิศทางเดียวกัน แสดงให้เห็นถึงประสิทธิภาพของกระบวนการข้างต้นได้เป็นอย่างดี ต่อมา Somprasertsri [94], นำเสนอแนวคิดการค้นคืนเอกสารด้วยกราฟความรู้ (Knowledge graphs) มีขั้นตอนสำคัญคือ 1) การประมวลผลข้อความ 2) การทำดัชนี 3) การวัดค่าความคล้ายคลึง และ 4) การจัดเรียงลำดับเอกสาร ค้นหาค่าความเหมือนด้วย Jaccard coefficient ร่วมกับ Tree graphs ผลการวิจัยสามารถนำมาใช้ในการค้นคืนเอกสารตามเนื้อหาที่เป็นประเด็นย่อยได้ดี โดยให้ค่า p เท่ากับ 0.75 ค่า r เท่ากับ 1.00 และค่า $F1$ เท่ากับ 0.84 ตามลำดับ

5.3 ระเบียบวิธีวิจัย

งานวิจัยในครั้งนี้ได้ดำเนินการศึกษารวบรวมข้อมูล แนวคิดและทฤษฎีที่เกี่ยวข้อง นำสู่การวิเคราะห์ข้อมูลด้านการท่องเที่ยว (Tourism domain) โดยชุมชนอีสานใต้ สำหรับวัตถุประสงค์ข้อที่ 3 โดยนำเสนอกระบวนการตามระเบียบวิธีวิจัยได้ 4 ประเด็นสำคัญดังนี้

5.3.1 คำถามวิจัย (Question research)

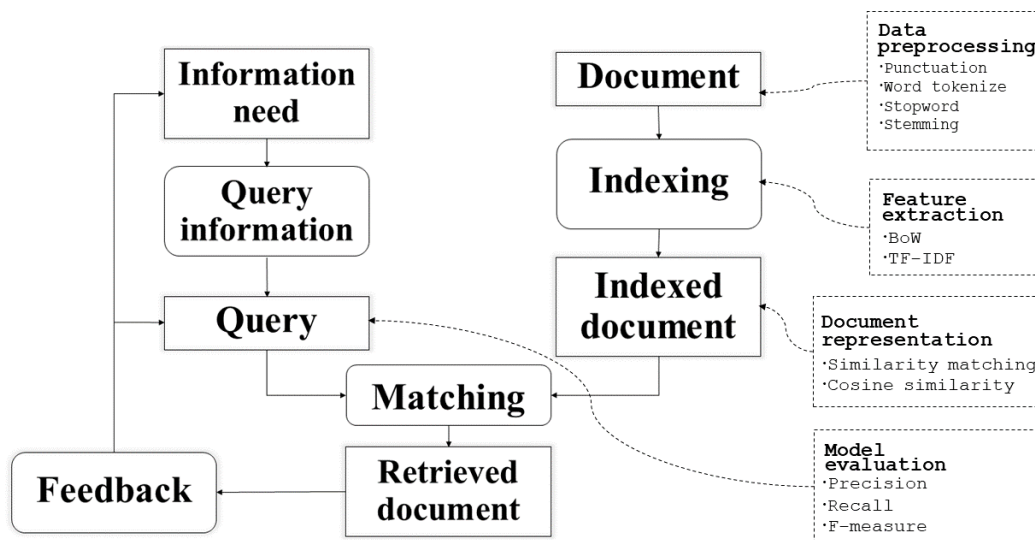
คำตอบสำหรับตอบคำถามอัตโนมัติสำหรับการท่องเที่ยวโดยชุมชนอีสานใต้ควรเป็นอย่างไร?

5.3.2 วัตถุประสงค์การวิจัย (Objective)

พัฒนากรอบการทำงานการหาคำตอบสำหรับตอบคำถามอัตโนมัติการท่องเที่ยวโดยชุมชนอีสานใต้

5.3.3 กรอบแนวคิดการวิจัย (Research framework)

งานวิจัยครั้งนี้มุ่งเน้นการค้นคืนสารสนเทศด้านการท่องเที่ยวชุมชนประเภทข้อความ (Textual information retrieval) จากความต้องการสารสนเทศ (Information need) [23] โดยรูปแบบการสอบถามสู่การสร้างกฎความสัมพันธ์ของนักท่องเที่ยวที่ปรากฏร่วมกันบ่อย (Frequent item set) การกำหนดและป้อนคำสอบถาม วิเคราะห์และนำมาเปรียบเทียบกับสำคัญที่อยู่ในดัชนี (Index) ซึ่งได้จากตัวแทนเอกสาร (Document representation) แล้วนำมาเปรียบเทียบ (Matching) กับเอกสารที่มีความสัมพันธ์กับคำสอบถามแล้วคืนผลการค้นหาให้กับผู้ใช้ หรือเอกสารที่เกี่ยวข้องกับความต้องการของผู้ใช้มากที่สุด แสดงดังรูปที่ 21



รูปที่ 21 กรอบแนวคิดการวิจัย (Conceptual framework)

ดำเนินการตามกรอบแนวคิดการวิจัย ทดลองและวิเคราะห์ข้อมูลบนพื้นฐานของโมเดลที่หลากหลาย โดยเลือกนำจุดเด่นของแต่ละโมเดลมาประยุกต์ใช้ สามารถอธิบายรายละเอียดพอสังเขปได้ 5 ขั้นตอนดังนี้

ขั้นตอนที่ 1 การเตรียมข้อมูล (Data preprocessing) ประกอบด้วย 4 โมเดลพื้นฐาน คือ 1) Text cleaning and text processing, 2) Word segmentation บนพื้นฐานของ Maximal matching (newmm) [43] ด้วยหลักการอ้างอิง Dictionary-based approach (DCB), 3) Stop word removal และ 4) Stemming เพื่อเพิ่มประสิทธิภาพในการค้นคืนด้วย Porter algorithm [44] [45].

ขั้นตอนที่ 2 การสร้างดัชนี (Indexing) หรือ การสกัดคุณลักษณะ (Feature extraction) ด้วยการนำความถี่มาคำนวณหาค่าน้ำหนักของคำ เพื่อสร้างเป็นตัวแทนเอกสารด้วยค่าน้ำหนักของคำ ซึ่งทั้งหมดจะถูกจัดเก็บอย่างมีมาตรฐานในฐานข้อมูล (Corpus) ผลลัพธ์ในขั้นตอนนี้ คือ 1) Dictionary, 2) Posting, and 3) Pointer.

ขั้นตอนที่ 3 ดัชนีเอกสาร (Indexed document) หรือตัวแทนเอกสาร (Document representation) ที่ได้จากขั้นตอนที่ 2 ผลลัพธ์ที่ได้รับในขั้นตอนนี้คือ 1) BoW และ 2) TF-IDF ซึ่งให้ความสำคัญกับความถี่ของคำ (Term frequency) และ น้ำหนักของคำ (term weight) ที่พบในเอกสารทั้งหมด เพื่อคำนวณหาความเหมือนของเอกสารทั้งคู่นบนพื้นฐาน VSM [18] ในขั้นตอนต่อไป

ขั้นตอนที่ 4 การเปรียบเทียบ (Matching) ด้วยการเปรียบเทียบความคล้ายคลึง (Similarity matching) ระหว่างคำสอบถามกับเอกสารคำตอบ โดยให้ความสำคัญกับ Term-document matrix ด้วยวิธีการทางสถิติเพื่อใช้วัดค่า Cosine similarity หรือ Distance ของเวกเตอร์ระหว่างเอกสารทั้งคู่

ดังนั้นค่า Cosine จะมีค่าอยู่ระหว่าง 0 ถึง 1 โดยผลลัพธ์ที่ได้จะถูกจัดเรียงจากค่าและความเหมือนตามลำดับ

ขั้นตอนที่ 5 การประเมินประสิทธิภาพโมเดล (Model evaluation) เป็นขั้นตอนสุดท้ายเพื่อสร้างความเชื่อมั่นก่อนนำโมเดลไปใช้งานจริง โดยความสามารถในการดึงเอาเอกสารที่มีความเกี่ยวข้อง (Relevant) กับคำถามแยกออกจากเอกสารที่ไม่เกี่ยวข้อง (Non-relevant) ซึ่งผลลัพธ์จะถูกจัดเรียงตามลำดับ โดยวิธีการประเมินประสิทธิภาพโมเดลในครั้งนี้ ผู้วิจัยดำเนินการใน 2 ส่วน คือ 1) Expert และ 2) System

5.3.4 การเก็บรวบรวมข้อมูล (Data collections)

จากการลงพื้นที่อย่างต่อเนื่องทำให้ผู้วิจัยได้ทราบถึงปัญหาและอุปสรรคที่ส่งผลกระทบต่อการทำงาน รวมถึงมุมมองด้านโอกาสในการส่งเสริมภาพลักษณ์และการตัดสินใจเดินทางของนักท่องเที่ยวในเขตอีสานใต้สู่การท่องเที่ยวและใช้ประโยชน์ในเชิงวิชาการบนพื้นฐานของงานวิจัยเดิมที่ถูกเผยแพร่ [22] [23] ใน 4 ด้านประกอบด้วย 1) การบริการที่เป็นเลิศ 2) สิ่งอำนวยความสะดวกที่มีมาตรฐาน 3) ข้อมูลสารสนเทศที่ถูกต้องและตรงกับความต้องการ และ 4) ประเพณีและวัฒนธรรมที่โดดเด่น เป็นข้อมูลนำเข้าด้วยข้อความภาษาไทย ทั้งหมด สามารถแบ่งข้อมูลนำเข้าเป็น 2 ชุด ได้แก่ ชุดคำถามและชุดคำตอบนำเสนอรายละเอียดดังต่อไปนี้

1) ชุดคำถาม (Query collections) เป็นข้อมูลนำเข้าที่ได้ผลลัพธ์จากวัตถุประสงค์ที่ 2 มี 3 ขั้นตอนดังนี้

ขั้นตอนที่ 1 วิเคราะห์รายการคำถามด้วยการวิเคราะห์หองค์ประกอบเชิงสำรวจ โดยวิธีการวิเคราะห์หองค์ประกอบหลัก เพื่อลดขนาดของตัวแปรและเลือกใช้วิธีหมุนแกนแบบตั้งฉาก (Varimax orthogonal rotation) ผลลัพธ์ที่ได้คือกลุ่มรายการคำถามที่มีความสัมพันธ์ไปในทิศทางเดียวกันแล้วนำมากำหนดชื่อกลุ่มใหม่คือ ESAN model ประกอบด้วย 4 กลุ่ม คือ 1) การบริการที่เป็นเลิศ 2) สิ่งอำนวยความสะดวกที่ได้มาตรฐาน 3) ข้อมูลสารสนเทศและการประชาสัมพันธ์ และ 4) ประเพณีวัฒนธรรมที่โดดเด่น ซึ่งการทดลองในขั้นตอนนี้ถูกเผยแพร่ในงานประชุมวิชาการระดับนานาชาติ [22] ข้อมูลสำหรับชุดคำถาม (Query collections) แสดงดังรูปที่ 22

Cluster E

- Rule①:** แหล่งท่องเที่ยวชุมชนมีห้องน้ำสะอาดมาตรฐานเพียงพอร้านค้าสินค้าบริการหลากหลาย และเจ้าหน้าที่ดูแลความปลอดภัยนักท่องเที่ยว
- Rule②:** แหล่งท่องเที่ยวชุมชนมีร้านค้ากำหนดราคาสินค้าบริการเหมาะสมหลากหลายเพียงพอ และเจ้าหน้าที่ดูแลความปลอดภัยนักท่องเที่ยว
- Rule③:** แหล่งท่องเที่ยวชุมชนมีร้านอาหารสะอาดได้มาตรฐานเพียงพอร้านค้าสินค้าบริการหลากหลาย และเจ้าหน้าที่ดูแลความปลอดภัยนักท่องเที่ยว
- Rule④:** แหล่งท่องเที่ยวชุมชนมีอาหารเอกลักษณ์บริการร้านค้าสินค้าหลากหลายเพียงพอและ เจ้าหน้าที่ดูแลความปลอดภัยนักท่องเที่ยว

Cluster S

- Rule①:** แหล่งท่องเที่ยวชุมชนมีกิจกรรมสะท้อนเอกลักษณ์วัฒนธรรมวิถีชีวิตภูมิปัญญาท้องถิ่น และที่พักปลอดภัย
- Rule②:** แหล่งท่องเที่ยวชุมชนมีโรงแรมที่พักปลอดภัยและกิจกรรมสะท้อนเอกลักษณ์วัฒนธรรมวิถีชีวิตภูมิปัญญาท้องถิ่น
- Rule③:** แหล่งท่องเที่ยวชุมชนมีกิจกรรมท่องเที่ยวความรู้ความเข้าใจเอกลักษณ์ชุมชนและโรงแรมที่พักปลอดภัย
- Rule④:** แหล่งท่องเที่ยวชุมชนมีกิจกรรมท่องเที่ยวความรู้ความเข้าใจเอกลักษณ์ชุมชนและ กิจกรรมสะท้อนวัฒนธรรมวิถีชีวิตภูมิปัญญาท้องถิ่น

45

Cluster A

- Rule①:** แหล่งท่องเที่ยวชุมชนแจ้งระเบียบข้อปฏิบัติมีระบบสารสนเทศเว็บไซต์ข้อมูล ประชาสัมพันธ์กิจกรรมการท่องเที่ยวช่องทางสังคมออนไลน์มีคู่มือเอกสารเผยแพร่มี เจ้าหน้าที่ที่แนะนำบริการมีคุณภาพกับนักท่องเที่ยว
- Rule②:** แหล่งท่องเที่ยวชุมชนแจ้งระเบียบข้อปฏิบัติมีระบบสารสนเทศเว็บไซต์ข้อมูลช่องทาง การสื่อสารแฟนเพจเฟสบุคไลน์คู่มือเอกสารเผยแพร่มีเจ้าหน้าที่แนะนำบริการมีคุณภาพกับ นักท่องเที่ยว
- Rule③:** แหล่งท่องเที่ยวชุมชนแจ้งระเบียบข้อปฏิบัติมีระบบสารสนเทศเว็บไซต์ข้อมูลช่องทาง การสื่อสารแฟนเพจเฟสบุคไลน์ประชาสัมพันธ์กิจกรรมการท่องเที่ยวช่องทางสังคม ออนไลน์มีเจ้าหน้าที่แนะนำบริการมีคุณภาพกับนักท่องเที่ยว

Cluster N

- Rule①:** ชุมชนมีแหล่งท่องเที่ยวหลากหลายอัตลักษณ์เฉพาะท้องถิ่นเอกลักษณ์โดดเด่นอนุรักษ์ วัฒนธรรมท้องถิ่น ประวัติศาสตร์วัฒนธรรมวิถีชีวิตและประเพณี
- Rule②:** ชุมชนมีแหล่งท่องเที่ยวอัตลักษณ์เฉพาะท้องถิ่นเอกลักษณ์โดดเด่นอนุรักษ์วัฒนธรรม ท้องถิ่นประวัติศาสตร์วัฒนธรรม วิถีชีวิตและประเพณี
- Rule③:** ชุมชนมีแหล่งท่องเที่ยวหลากหลายอัตลักษณ์เฉพาะท้องถิ่นอนุรักษ์วัฒนธรรมท้องถิ่น ประวัติศาสตร์วัฒนธรรม วิถีชีวิตและประเพณี

46

รูปที่ 22 การเตรียมข้อมูลสำหรับชุดคำถาม (Query collections)

ขั้นตอนที่ 2 ค้นหากฎความสัมพันธ์ ด้วย Apriori algorithm จากกลุ่มรายการคำถามที่ได้จากขั้นตอนแรก โดยพิจารณาเลือกจาก 2 ส่วน คือ 1) ค่า Confidence ตั้งแต่ 98% ขึ้นไป และ 2) กฎที่มีลักษณะเด่นในเชิงประจักษ์สอดคล้องกับแนวคิด Maslow's Hierarchy of needs โดยตัดกฎที่มีความซ้ำซ้อนทิ้ง [40] สามารถสร้างกฎที่ดีที่สุด (Best rules) ได้ 14 กฎ และผลการทดลองในครั้งนี้ได้ถูกตีพิมพ์และเผยแพร่ในวารสารระดับนานาชาติ [23]

ขั้นตอนที่ 3 กระบวนการ Data pre-processing มี 4 ขั้นตอนสำคัญ คือ 1) การทำความสะอาดข้อมูล (Text cleaning) ด้วยการลบอักขระพิเศษและช่องว่าง (Space and punctuation removal) 2) การตัดคำ (Word segmentation หรือ Tokenize) 3) การลบคำหยุด หรือ คำที่ไม่สื่อความหมาย (Stop word removal) และ 4) การหารากศัพท์ (Stemming)

2) ชุดเอกสารคำตอบ (Answer collections) ดำเนินการรวบรวมข้อมูลด้านการท่องเที่ยวอีสานได้ใน 4 จังหวัด ประกอบด้วย นครราชสีมา บุรีรัมย์ สุรินทร์ และศรีสะเกษ รวมเอกสารทั้งสิ้น 1195 ไฟล์เอกสาร ซึ่งเป็นข้อมูลปฐมภูมิและทุติยภูมิ อาทิเช่น Search engine เอกสาร แผ่นพับ และไฟล์ดิจิทัล จากหน่วยงานที่เกี่ยวข้อง จัดเก็บข้อมูลในรูปแบบไฟล์เอกสารด้วยโปรแกรมไมโครซอฟท์เวิร์ด (แสดงดังตารางที่ 10) เพื่อนำทั้งหมดเข้าสู่ขั้นตอนการเตรียมข้อมูล (Data pre-processing) คล้ายกับการเตรียมข้อมูลในชุดคำถาม (Query collections) โดยเพิ่มการหารากศัพท์ (Stemming) ด้วยการแปลงรูปแบบของคำที่แตกต่างกันให้อยู่ในรูปแบบที่คล้ายกัน

ตารางที่ 10 แสดงรายการชุดเอกสารคำตอบ (Answer collections) ทั้ง 4 กลุ่ม (Cluster)

กลุ่มเอกสาร (Class/Label)	ฐานข้อมูล (Corpus)		
	เอกสาร (ไฟล์)	สัดส่วน (%)	หน่วยความจำ (MB.)
การบริการที่เป็นเลิศ	524	42.36	8.54
สิ่งอำนวยความสะดวกที่ได้มาตรฐาน	267	21.58	4.3
ข้อมูลสารสนเทศและประชาสัมพันธ์	201	16.25	3.23
ประเพณีวัฒนธรรมที่โดดเด่น	245	19.81	3.78
รวม	1237	100.00	19.85

จากนั้นนำข้อมูลทั้งหมดเข้าสู่ขั้นตอนการเตรียมข้อมูล (Data preprocessing) ด้วยการทำความสะอาดข้อมูล (Text cleaning) ด้วยการลบช่องว่างและสัญลักษณ์พิเศษ (Space and punctuation removal) เหมือนกับการเตรียมข้อมูลในชุดคำถาม ประกอบด้วย 3 ขั้นตอนสำคัญ

คือ 1) การตัดคำ (Word segmentation) 2) การลบคำหยุด หรือ คำที่ไม่สื่อความหมาย (Stop word removal) และ 3) การหารากศัพท์ (Stemming) ผลลัพธ์ที่ได้คือ คำ (Term) ถุงคำ (Bag of word: BoW) และค่าน้ำหนักคำ (Term vector) เพื่อใช้เป็นชุดคำตอบจากคำตอบถามของผู้ใช้

5.4 ผลการทดลอง

การวิจัยในขั้นตอนนี้ มุ่งเน้นสร้างต้นแบบการค้นคืนสารสนเทศประเภทข้อความด้านการท่องเที่ยว กรณีศึกษาชุมชนอีสานได้ เพื่อใช้เป็นแนวทางแก้ปัญหาการเข้าถึงข้อมูลของนักท่องเที่ยวและผู้มีส่วนเกี่ยวข้อง โดยนักวิจัยเลือกใช้ PyThaiNLP สำหรับการประมวลผลข้อมูลภาษาไทยด้วยโปรแกรมภาษา Python ซึ่งเป็นเครื่องมือที่ให้บริการโดย Google colabatory [77] โดยนำเสนอผลการทดลองและการวิเคราะห์ข้อมูลใน 5 ขั้นตอนดังนี้

ขั้นตอนที่ 1 การเตรียมข้อมูล (Data preprocessing) เริ่มต้นด้วยการทำความสะอาดข้อมูล (Text cleaning) ด้วยการลบช่องว่างและสัญลักษณ์พิเศษ (Space and punctuation removal) ดำเนินการบน 3 โมเดลพื้นฐาน คือ 1) การตัดคำ (Word segmentation) บนพื้นฐานของ Maximal matching (newmm) [30] ด้วยหลักการอ้างอิง Dictionary-based approach (DCB) 2) การลบคำหยุด หรือ คำที่ไม่สื่อความหมาย (Stop word removal) และ 3) Stemming แปลงรูปแบบของคำที่แตกต่างกันให้อยู่ในรูปแบบที่คล้ายกัน เพื่อเพิ่มโอกาสในการค้นคืนด้วย Porter algorithm [27] [28] โดยนำเสนอผลการทดลองดังรูปที่ 23

Column2
0 [องค์, การบริหาร, การพัฒนา, พื้นที่, พิเศษ, การท่องเที่ยว, ยั่งยืน, 'ก', 'ก', การประชุม, '...', 'รับฟัง, 'ความคิดเห็น, 'ภาคี, 'เครือข่าย, 'การท่องเที่ยว, 'ประกาศ, 'พื้นที่, 'พื้น
1 [ไฮไลต์, 'อีสาน, 'จังหวัด, 'สถานที่ท่องเที่ยว, 'หยุด, 'ยอดเยี่ยม, 'โควิด, 'ก', 'ก', 'ไทยรัฐ, 'ออนไลน์, 'ไลฟ์สไตล์, 'ท่องเที่ยว, 'ก', 'ก', 'หลังจากที่, 'สถานที่ท่องเที่ยว, 'จังหวัด
2 [ท่องเที่ยว, 'เอช, 'เจ้า, 'ต้นกล้วย, 'ก', 'ก', 'แผน, 'ว่าด้วย, 'เมือง, 'เรื่อง, 'การท่องเที่ยว, 'ก', 'ก', 'เป้าหมาย, 'การท่องเที่ยว, 'แผน, 'พัฒนา, 'เศรษฐกิจ, 'สังคม, 'แห่งชาติ, 'จ
3 [พช, 'ปลื้ม, 'แสน, 'คน, 'แห่, 'ชม, 'งาน, 'ชมชน, 'ท่องเที่ยว, 'otop, 'น', 'วัด, 'วิถี, 'ฮี, 'เวิน, 'ดี, 'เสริม, 'ชมชน, 'สี, 'โควิด, 'ก', 'ก', 'กรมการพัฒนาชุมชน, 'ภารกิจ, 'ส่งเสริม
4 [พาณิชย์, 'หนุน, 'ผู้ประกอบการ, 'บุรีรัมย์, 'ก', 'ก', 'ต่อยอด, 'อัตลักษณ์, 'วัฒนธรรม, 'เครือข่ายระดับ, 'ชน, 'เผ่า, 'อีสาน, 'เตรียม, 'โก, 'อินเตอร์, 'เข้าสู, 'สากล, 'ก', 'ก', 'พาณิชย์
5 [เที่ยว, 'บุรีรัมย์, 'อำเภอ, 'กระสัง, 'ศูนย์, 'วัฒนธรรม, 'อีสาน, 'ก', 'ก', 'ศูนย์, 'วัฒนธรรม, 'อีสาน, 'กระสัง, 'บุรีรัมย์, 'ก', 'ก', 'ศูนย์, 'วัฒนธรรม, 'อีสาน, 'ตั้งอยู่, 'ด้านหน้า, 'น
6 [อพท, 'เวที, 'ความ, 'ยกระดับ, 'จังหวัด, 'อีสาน, 'เข้าเกณฑ์, 'gstc, 'ก', 'ก', 'อพท, 'ผนัก, 'เครือข่าย, 'จังหวัด, 'อีสาน, 'เตรียม, 'พัฒนา, 'พื้นที่, 'การท่องเที่ยว, 'ยั่งยืน, 'ประ
7 [อพท, 'ผนัก, 'ภาคี, 'เครือข่าย, 'จังหวัด, 'อีสาน, 'ลุย, 'พัฒนา, 'พื้นที่, 'การท่องเที่ยว, 'ยั่งยืน, 'ก', 'ก', 'ช, 'วิทย์, 'มิตร, 'ซ่อม, 'รักษาการ, 'ผู้อำนวยการ, 'องค์, 'การบริหาร, '
8 [รวม, 'ท่องเที่ยว, 'ขับเคลื่อน, 'อารยธรรม, 'อีสาน, 'เชื่อมโยง, 'ท่องเที่ยว, 'เชิง, 'บูรณาการ, 'ภูมิภาค, 'ก', 'ก', 'พื้นที่, 'กัญชามา, 'ห้อง, 'ศรี, 'พิมาน, 'โรงแรม, 'สุรินทร์, 'เจ้า, 'ส
9 [ทริป, 'บุรีรัมย์, 'สุรินทร์, 'ศรีสะเกษ, 'อุบลฯ, 'อุ, 'ช้าง, 'อุ, 'กระเรียน, 'ก', 'ก', 'เดินทาง, 'วันที่, 'ละ, 'คืน, 'บาท, 'ท่าน, 'จำนวน, 'ผู้เดินทาง, 'ต่ำกว่า, 'ท่าน, 'ยานพาหนะ, 'รอ
10 [องค์, 'การบริหาร, 'การพัฒนา, 'พื้นที่, 'พิเศษ, 'การท่องเที่ยว, 'ยั่งยืน, 'องค์การมหาชน, 'ก', 'ก', 'บุรีรัมย์, 'นำร่อง, 'อารย, 'สถาปัตย์, 'ยกระดับ, 'อีสาน, 'แหล่งท่องเที่ยว, 'คน
11 [องค์, 'การบริหาร, 'การพัฒนา, 'พื้นที่, 'พิเศษ, 'การท่องเที่ยว, 'ยั่งยืน, 'องค์การมหาชน, 'ก', 'ก', 'อพท, 'เตรียม, 'ประกาศ, '...', 'บุรีรัมย์, '...', 'พื้นที่, 'พิเศษ, 'ก', 'ก', 'อพท, 'เ
12 [องค์, 'การบริหาร, 'การพัฒนา, 'พื้นที่, 'พิเศษ, 'การท่องเที่ยว, 'ยั่งยืน, 'องค์การมหาชน, 'ก', 'ก', 'อีสาน, 'ดินแดน, 'ความคิดสร้างสรรค์, 'ภูมิปัญญา, 'วัฒนธรรม, 'ก', 'ก', 'อพท
13 [เที่ยว, 'อีสาน, 'อุ, 'ล, 'หลวง, 'แม่ดิน, 'ถิ่น, 'อีสาน, 'isan, 'ก', 'love, 'ก', 'ก', 'ทพ, 'ท, 'อีสาน, 'ชวน, 'สัมผัส, 'อัตลักษณ์, 'อุ, 'ล, 'อีสาน, 'ใคร่, 'การเมือง, 'รอง, 'ขวัญ
14 [สุ, 'วังน, '...', 'ฉลอง, 'ปี, 'ตรุษจีน, 'โคราช, 'สาม, 'สัมพันธ์, 'ไทย, 'จีน, 'กระตุ้น, 'ศก, 'สร้าง, 'รายได้, 'ชมชน, 'ก', 'ก', 'กระหม, 'ตรุษจีน, 'โคราช, '...', 'สุ, 'วังน, '...', 'จับมือ, 'เ
15 [ผู้ว่า, 'เดินทาง, 'เมือง, 'โคราช, 'นำอยู่, 'เร่ง, 'คืน, 'ถนน, 'sang, 'ปัญหา, 'อุ, 'ไฟฟ้า, 'ยักษ์, 'โผล่, 'ทางเท้า, 'เก็ล่อน, 'เมือง, 'ก', 'ก', 'ศูนย์ข่าว, 'บรรณานิคม, 'ผู้ว่า, 'โคร
16 [หน้า, 'หลัก, 'ภูมิภาค, 'ตะวันออกเฉียงเหนือ, 'ก', 'ก', 'นักท่องเที่ยว, 'เฮ, 'ชม, 'นศกรชัย, 'ดวงอาทิตย์, 'ดก, 'ช่อ, 'ประดู่, '...', 'พนม, 'รุ่ง, '...', 'ชัดเจน, 'ก', 'ก', 'ก', 'ก', 'ก', 'ก
17 [หน้า, 'หลัก, 'ภูมิภาค, 'ตะวันออกเฉียงเหนือ, 'ก', 'ก', 'บุรีรัมย์, '...', 'เทศกาล, 'ข่าว, 'หอม, 'จ่อม, 'กึ่ง, 'ชม, 'ทุ่ง, 'ก', 'ประโคน, 'ชัย, '...', 'ปีใหม่, 'ก', 'ก', 'ก', 'ก', 'ก', 'ก', 'ก', 'ก', 'ก', 'ก
18 [หน้า, 'หลัก, 'ภูมิภาค, 'ตะวันออกเฉียงเหนือ, 'ก', 'ก', 'หอ, 'คำ, 'บุรีรัมย์, 'ชี, 'เงิน, 'สะพัด, 'ปีใหม่, 'ล้าน, 'แห่, 'เที่ยว, 'พนม, 'รุ่ง, 'ปราสาท, 'เมือง, 'คำ, 'ก', 'ก', 'ก', 'ก', 'ก', 'ก', 'ก', 'ก', 'ก', 'ก
19 [เที่ยว, 'โคราช, 'ดินแดน, 'ที่ราบสูง, 'ประดู่, 'ภาคอีสาน, 'ก', 'ก', 'เที่ยว, 'โคราช, 'ดินแดน, 'ที่ราบสูง, 'แหล่งท่องเที่ยว, 'เมือง, 'อุดมสมบูรณ์, 'ทรัพยากร, 'สิ่งอำนวยความสะดวก
20 [ทพ, 'ท, 'โคราช, 'ชวน, 'เที่ยว, 'วิถี, 'ปลอดภัย, 'มาตรฐาน, 'sha, 'พื้นที่, 'เศรษฐกิจ, 'การท่องเที่ยว, 'ก', 'ก', 'ก', 'ก', 'ก', 'ทพ, 'ท, 'โคราช, 'ชวน, 'คน, 'โคราช, 'เที่ยว, 'โคราช

รูปที่ 23 ผลการทดลองการเตรียมข้อมูล (Data preprocessing)

1) การตัดคำ (Tokenization) คือการทำให้เอกสารถูกจัดอยู่ในรูปแบบชุดของตัวอักษร (String) หรือ การแยกข้อความออกเป็นหน่วยย่อย ๆ เพื่อหาขอบเขตของคำเป็นหน่วยภาษาที่เล็กที่สุด แบ่งได้ 3 กลุ่มหลัก คือ 1) Rule-based Approach 2) Corpus-based approach และ 3)

Dictionary-based approach (DCB) ซึ่งวิธีอ้างอิงจากพจนานุกรมสามารถสื่อความหมายได้บนพื้นฐานของการตัดคำแบบสอดคล้องมากที่สุด (Maximal matching: newmm) [44] [45] [43] จากเหตุผลข้างต้นการวิจัยในครั้งนี้จึงเลือกวิธี newmm

2) การลบคำหยุด (Stop words) ค้นหาที่ไม่มีประโยชน์ออกโดยที่ความหมายของเอกสารไม่เปลี่ยนแปลงจากความหมายเดิม คำฟุ่มเฟือยในภาษาไทยที่พบบ่อย อาทิเช่น “มี” (have), “โดย” (by), “คือ” (is), “เพื่อ” (for), and “ซึ่ง” (then) โดยจุดเด่นของการทำ Stop words จะส่งผลดีทำให้คลังคำศัพท์ (Bag of words: BoW) ไม่ใหญ่เกินไปโดยไม่จำเป็น ผลลัพธ์ที่ได้จากกระบวนการนี้คือ ลดคำที่ไม่สำคัญออกจากขั้นตอนการสกัดคุณลักษณะ (Feature extraction)

3) การหารากศัพท์ (Stemming words) เป็นการหารากศัพท์ด้วยการแปลงรูปแบบของคำที่แตกต่างกันให้อยู่ในรูปแบบที่คล้ายกัน อาทิเช่น “State (รัฐ)” และ “City (เมือง)” อ้างอิงจากพจนานุกรม (Dictionary-based: DCB) เพื่อให้การอธิบายเป็นไปในทางเดียวกัน การหารากศัพท์จะดำเนินการก่อนการทำดัชนี (Indexing) เพื่อเพิ่มประสิทธิภาพการค้นคืน อาทิเช่น Porter algorithm

ขั้นตอนที่ 2 การสร้างดัชนี (Indexing) หรือ การสกัดคุณลักษณะ (Feature extraction) ด้วยการนำความถี่มาคำนวณหาค่าน้ำหนักของคำ เพื่อสร้างเป็นตัวแทนเอกสารด้วยค่าน้ำหนักของคำ ซึ่งทั้งหมดจะถูกจัดเก็บอย่างมีมาตรฐานในฐานข้อมูล (Corpus) ผลลัพธ์ในขั้นตอนนี้ คือ 1) Dictionary 2) Posting และ 3) Pointer กระบวนการสกัดคุณลักษณะเอกสาร แสดงดังรูปที่ 24

Token		Term document		Term sorting		Term document matrix				
docID	documents	term	docID	term	docID	Term	doc1	doc2	doc3	doc4
doc1	[แหล่งท่องเที่ยว, มี, สิ่งอำนวยความสะดวก, พื้นฐาน, เพียงพอ]	แหล่งท่องเที่ยว	1	การจัด	4	การจัด	0	0	0	1
		มี	1	จราจร	4	จราจร	0	0	0	1
		สิ่งอำนวยความสะดวก	1	จอดรถ	3	จอดรถ	0	0	1	0
		พื้นฐาน	1	ด้านสุขภาพ	2	ด้านสุขภาพ	0	1	0	0
doc2	[แหล่งท่องเที่ยว, มี, สถานบริการ, ด้านสุขภาพ]	แหล่งท่องเที่ยว	2	ได้	4	ได้	0	0	0	1
		มี	2	ที่เหมาะสม	3	ที่เหมาะสม	0	0	1	0
		สถานบริการ	2	ปลอดภัย	3	ปลอดภัย	0	0	1	0
		ด้านสุขภาพ	2	พื้นฐาน	1	พื้นฐาน	1	0	0	0
doc3	[แหล่งท่องเที่ยว, มี, สถานที่, จอดรถ, เหมาะสม, และ, ปลอดภัย]	แหล่งท่องเที่ยว	3	มี	1	มี	1	0	0	0
		มี	3	มี	2	เพียงพอ	1	0	0	0
		สถานที่	3	มี	3	มี	1	1	1	1
		จอดรถ	3	มี	4	ระเบียบ	0	0	0	1
doc4	[แหล่งท่องเที่ยว, มี, การจัด, ระเบียบ, และ, จราจร, ได้, อย่างเหมาะสม]	ที่เหมาะสม	3	และ	3	และ	0	0	1	1
		ปลอดภัย	3	และ	4	สถานที่	0	0	1	0
		แหล่งท่องเที่ยว	4	สถานี่	3	สถานบริการ	0	1	0	0
		มี	4	สถานบริการ	2	สิ่งอำนวยความสะดวก	1	0	0	0
		การจัด	4	สิ่งอำนวยความสะดวก	1	แหล่งท่องเที่ยว	1	1	1	1
		ระเบียบ	4	แหล่งท่องเที่ยว	1	แหล่งท่องเที่ยว	1	1	1	1
		จราจร	4	แหล่งท่องเที่ยว	2	แหล่งท่องเที่ยว	1	1	1	1
		ได้	4	แหล่งท่องเที่ยว	3	แหล่งท่องเที่ยว	1	1	1	1
อย่างเหมาะสม	4	อย่างเหมาะสม	4	อย่างเหมาะสม	0	0	0	1		

รูปที่ 24 ตัวอย่างกระบวนการสกัดคุณลักษณะเอกสาร

ขั้นตอนที่ 3 ดัชนีเอกสาร (Indexed document) ที่ได้จากขั้นตอนที่ 2 หรือตัวแทนเอกสาร (Document representation) ที่ให้ความสำคัญกับความถี่ของคำ (Term frequency) และ น้ำหนักของคำ (Term weight) ที่พบในเอกสารทั้งหมด เพื่อคำนวณหาความเหมือนของเอกสารทั้งฉบับพื้นฐาน ปริภูมิเวกเตอร์ (Vector space model: VSM) [21] ในขั้นตอนต่อไป โดยผลลัพธ์ที่ได้รับในขั้นตอนนี้คือ 1) ถูงคำ (Bag of words: BoW) และ 2) TF-IDF แสดงดังตารางที่ 11

ตารางที่ 11 แสดงตัวอย่างเวกเตอร์ค่าน้ำหนักคำ (Term vector) ทั้ง 4 กลุ่ม (Cluster)

Class	IDF score	TF-IDF score
Excellent service	{'แหล่งท่องเที่ยว': 1.14568, 'ห้องน้ำ': 1.74027, 'สะอาด': 1.65780, 'มาตรฐาน': 1.66428, 'นักท่องเที่ยว': 1.0, ...'ชุมชน': 1.07559}	{'แหล่งท่องเที่ยว': 0.00703, 'ห้องน้ำ': 0.01068, 'สะอาด': 0.010178, 'มาตรฐาน': 0.02042, 'นักท่องเที่ยว': 0.00613, ...'ชุมชน': 0.00660}
Standard facilities	{'กิจกรรม': 1.51923, 'การท่องเที่ยว': 1.0, 'เอกลักษณ์': 1.12745, 'วัฒนธรรม': 1.39630, ...'ชุมชน': 1.07559}	{'กิจกรรม': 0.01356, 'การท่องเที่ยว': 0.00893, 'เอกลักษณ์': 0.01007, 'วัฒนธรรม': 0.01247, ...'ชุมชน': 0.00960}
Accuracy of information	{'แหล่งท่องเที่ยว': 1.14568, 'ระเบียบ': 1.0, 'ข้อปฏิบัติ': 2.77270, 'นักท่องเที่ยว': 1.0, ... 'การสื่อสาร': 1.0}	{'แหล่งท่องเที่ยว': 0.00606, 'ระเบียบ': 0.00529, 'ข้อปฏิบัติ': 0.01467, 'นักท่องเที่ยว': 0.00529, 'สารสนเทศ': ... 'การสื่อสาร': 0.00529}
Noble culture	{'ชุมชน': 1.07559, 'แหล่งท่องเที่ยว': 1.14568, 'หลากหลาย': 1.86930, 'เอกลักษณ์': 2.39554, ... 'ประเพณี': 2.09194}	{'ชุมชน': 0.00717, 'แหล่งท่องเที่ยว': 0.00764, 'หลากหลาย': 0.01246, 'เอกลักษณ์': 0.01597, ... 'ประเพณี': 0.01395}

ขั้นตอนที่ 4 การเปรียบเทียบ (Matching) ด้วยการเปรียบเทียบความคล้ายคลึง (Similarity matching) ระหว่างคำสอบถามกับเอกสารคำตอบ โดยให้ความสำคัญกับ Term-document matrix ด้วยวิธีการทางสถิติเพื่อใช้วัดค่าความคล้ายคลึงแบบโคไซน์ (Cosine similarity) หรือเรียกได้ว่าเป็นการวัดระยะห่าง (Distance) ของเวกเตอร์ระหว่างเอกสารทั้งคู่ ดังนั้นค่าโคไซน์ จะมีค่าอยู่ระหว่าง 0 ถึง 1 โดย 0 คือมีความเหมือนกันน้อยที่สุด และ 1 คือมีความเหมือนกันมากที่สุด แสดงผลการเปรียบเทียบความเหมือนระหว่างคำสอบถามและเอกสารคำตอบแล้วนำตัวเลขมาคำนวณด้วยวิธีการทางสถิติ Cosine similarity [84] [83] รวมทั้งสิ้น 4 กลุ่ม แสดงดังตารางที่ 12-15

ตารางที่ 12 แสดงค่าความคล้ายคลึง 3 ลำดับแรก กลุ่ม E ด้านการบริการที่เป็นเลิศ

คำสอบถาม (Query)	หมายเลขเอกสาร (Doc-id)	ค่าความคล้ายคลึง (Cosine similarity)
Rule1: แหล่งท่องเที่ยวชุมชนมีห้องน้ำสะอาดมาตรฐานเพียงพอ ร้านค้าสินค้าบริการหลากหลายและเจ้าหน้าที่ดูแลความปลอดภัย นักท่องเที่ยวมั้ย	325	0.892
	266	0.816
	289	0.795
Rule2: แหล่งท่องเที่ยวชุมชนมีร้านค้ากำหนดราคาสินค้าบริการ เหมาะสมหลากหลายเพียงพอและเจ้าหน้าที่ดูแลความปลอดภัย นักท่องเที่ยวมั้ย	181	0.863
	383	0.842
	160	0.810
Rule3: แหล่งท่องเที่ยวชุมชนมีร้านอาหารสะอาดได้มาตรฐาน เพียงพอร้านค้าสินค้าบริการหลากหลายและเจ้าหน้าที่ดูแลความ ปลอดภัยนักท่องเที่ยวมั้ย	452	0.874
	283	0.869
	144	0.858
Rule4: แหล่งท่องเที่ยวชุมชนมีอาหารเอกลักษณ์บริการร้านค้าสินค้า หลากหลายเพียงพอและเจ้าหน้าที่ดูแลความปลอดภัยนักท่องเที่ยวมั้ย	416	0.867
	72	0.806
	2	0.786

ตารางที่ 13 แสดงค่าความคล้ายคลึง 3 ลำดับแรก กลุ่ม S ด้านสิ่งอำนวยความสะดวกที่ได้มาตรฐาน

คำสอบถาม (Query)	หมายเลขเอกสาร (Doc-id)	ค่าความคล้ายคลึง (Cosine similarity)
Rule1: แหล่งท่องเที่ยวชุมชนมีกิจกรรมสะท้อนเอกลักษณ์วัฒนธรรมวิถีชีวิต ภูมิปัญญาท้องถิ่นและที่พักปลอดภัยมั้ย	64	0.872
	265	0.780
	101	0.766
Rule2: แหล่งท่องเที่ยวชุมชนมีโรงแรมที่พักปลอดภัยและกิจกรรมสะท้อน เอกลักษณ์วัฒนธรรมวิถีชีวิตภูมิปัญญาท้องถิ่นมั้ย	133	0.820
	36	0.864
	263	0.869
Rule3: แหล่งท่องเที่ยวชุมชนมีกิจกรรมท่องเที่ยวความรู้ความเข้าใจเอกลักษณ์ ชุมชนและโรงแรมที่พักปลอดภัยมั้ย	165	0.943
	164	0.895
	84	0.875
Rule4: แหล่งท่องเที่ยวชุมชนมีกิจกรรมท่องเที่ยวความรู้ความเข้าใจเอกลักษณ์ ชุมชนและกิจกรรมสะท้อนวัฒนธรรมวิถีชีวิตภูมิปัญญาท้องถิ่นมั้ย	135	0.958
	142	0.881
	236	0.878

ตารางที่ 14 แสดงค่าความคล้ายคลึง 3 ลำดับแรก กลุ่ม A ด้านข้อมูลสารสนเทศและประชาสัมพันธ์

คำสอบถาม (Query)	หมายเลขเอกสาร (Doc-id)	ค่าความคล้ายคลึง (Cosine similarity)
Rule1: แหล่งท่องเที่ยวชุมชนเชิงระเบียบข้อปฏิบัติมีระบบ	81	0.867
สารสนเทศเว็บไซต์ข้อมูลประชาสัมพันธ์กิจกรรมการท่องเที่ยว	143	0.813
ช่องทางสังคมออนไลน์มีคู่มือเอกสารเผยแพร่มีเจ้าหน้าที่แนะนำ		
บริการมีคุณภาพกับนักท่องเที่ยว	179	0.785
Rule2: แหล่งท่องเที่ยวชุมชนเชิงระเบียบข้อปฏิบัติมีระบบ	81	0.840
สารสนเทศเว็บไซต์ข้อมูลช่องทางการสื่อสารแฟนเพจเฟซบุ๊ก	32	0.792
คู่มือเอกสารเผยแพร่มีเจ้าหน้าที่แนะนำบริการมีคุณภาพกับ		
นักท่องเที่ยว	141	0.764
Rule3: แหล่งท่องเที่ยวชุมชนเชิงระเบียบข้อปฏิบัติมีระบบ	173	0.771
สารสนเทศเว็บไซต์ข้อมูลช่องทางการสื่อสารแฟนเพจเฟซบุ๊ก	114	0.741
ประชาสัมพันธ์กิจกรรมการท่องเที่ยวช่องทางสังคมออนไลน์มี		
เจ้าหน้าที่แนะนำบริการมีคุณภาพกับนักท่องเที่ยว	156	0.712

ตารางที่ 15 แสดงค่าความคล้ายคลึง 3 ลำดับแรก กลุ่ม N ด้านประเพณีวัฒนธรรมที่โดดเด่น

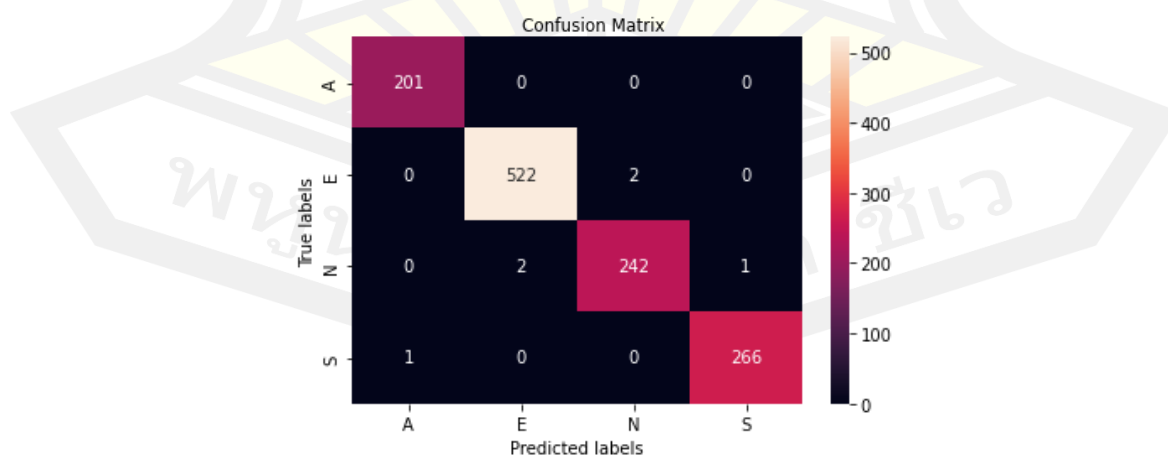
คำสอบถาม (Query)	หมายเลขเอกสาร (Doc-id)	ค่าความคล้ายคลึง (Cosine similarity)
Rule1: ชุมชนมีแหล่งท่องเที่ยวหลากหลายอัตลักษณ์เฉพาะท้องถิ่น	22	0.745
เอกลักษณ์โดดเด่นอนุรักษ์วัฒนธรรมท้องถิ่น ประวัติศาสตร์	122	0.751
วัฒนธรรมวิถีชีวิตและประเพณี	209	0.769
Rule2: ชุมชนมีแหล่งท่องเที่ยวอัตลักษณ์เฉพาะท้องถิ่นเอกลักษณ์	79	0.713
โดดเด่นอนุรักษ์วัฒนธรรมท้องถิ่นประวัติศาสตร์วัฒนธรรม วิถีชีวิต	22	0.723
และประเพณี	242	0.744
Rule3: ชุมชนมีแหล่งท่องเที่ยวหลากหลายอัตลักษณ์เฉพาะท้องถิ่น	122	0.805
อนุรักษ์วัฒนธรรมท้องถิ่นประวัติศาสตร์วัฒนธรรม วิถีชีวิตและ	72	0.796
ประเพณี	2	0.781

ขั้นตอนที่ 5 การประเมินประสิทธิภาพโมเดล (Model evaluation) เป็นขั้นตอนสุดท้ายเพื่อสร้างความเชื่อมั่นก่อนนำโมเดลไปใช้งานจริง โดยความสามารถในการดึงเอาเอกสารที่มีความเกี่ยวข้อง (Relevant) กับคำถามแยกออกจากเอกสารที่ไม่เกี่ยวข้อง (Non-relevant) ซึ่งผลลัพธ์จะถูกจัดเรียงตามลำดับ ดังตารางที่ 16

ตารางที่ 16 แสดงจำนวนเอกสารในกลุ่มเอกสาร E-N

กลุ่มเอกสาร (Class/Label)	ฐานข้อมูล (Corpus) Document
Class E	524
Class S	267
Class A	201
Class N	245
รวม	1237

จากตารางที่ 16 ผู้วิจัยได้แบ่งชุดข้อมูลสำหรับตรวจคำตอบหรือประเมินความถูกต้อง ด้วยการกำหนดคำตอบจากการวิเคราะห์ข้อมูลที่เรียกว่าคลาส (Class) หรือ ลาเบล (Label) การทดลองในครั้งนี้แบ่งชุดข้อมูล จำนวน 4 กลุ่ม รวมชุดข้อมูลทั้งสิ้น 1237 เอกสาร ประกอบด้วย Class E จำนวน 524 เอกสาร Class S จำนวน 267 เอกสาร Class A จำนวน 201 เอกสาร และ Class N จำนวน 245 เอกสาร เพื่ออธิบายคำตอบและประสิทธิภาพของโมเดลด้วยเมตริกซ์วัดประสิทธิภาพ (Confusion Matrix) ที่สร้างขึ้นด้วยแบบจำลองทางคณิตศาสตร์ แสดงดังรูปที่ 25 และตารางที่ 17



รูปที่ 25 เมตริกซ์วัดประสิทธิภาพ (Confusion Matrix) ชุดข้อมูลทั้ง 4 Class

ตารางที่ 17 แสดงผลการตรวจคำตอบหรือประเมินความถูกต้องด้วยเมตริกซ์วัดประสิทธิภาพ

Class	Confusion matrix				Model evaluation		
	Class E	Class S	Class A	Class N	<i>P</i>	<i>R</i>	<i>F1</i>
Class E	522	2	0	0	1	1	1
Class S	0	266	1	0	1	1	1
Class A	0	0	201	0	1	1	1
Class N	2	1	0	242	0.99	0.99	0.99
	Accuracy			0.99			
	Macro avg	1	0.99	1			
	Weighted avg	1	1	1			

จากตารางที่ 17 สามารถสรุปผลการประเมินโดยระบบ (System) เพื่ออธิบายคำตอบที่ได้จากการค้นคืนเอกสารและการประเมินประสิทธิภาพของชุดข้อมูลคำตอบ (Document/Answer) รวม 1,237 ไฟล์เอกสาร พบว่าให้ค่าความถูกต้องร้อยละ 99% โดยโมเดลที่สร้างขึ้นของทั้ง 4 คลาส (class) คือ 1) Class E 2) Class S 3) Class A และ 4) Class N รวมเอกสารทั้งสิ้น 1237 เอกสาร โดยผลการทำนายแบ่งได้ 2 กลุ่ม คือ กลุ่มที่ทำนายถูกต้อง และ ทำนายไม่ถูกต้อง ผลการทดลองพบว่า

Class E มีข้อมูลรวม 524 เอกสาร ทำนายถูกต้อง 522 เอกสาร และ ทำนายไม่ถูกต้องรวม 2 เอกสาร โดยทำนายเป็น Class S จำนวน 2 เอกสาร ผลการทดลองสามารถอธิบายคำตอบการค้นคืนเอกสารและให้ค่าความแม่นยำ (Precision) เท่ากับ 1 ให้ค่าระลึก (Recall) เท่ากับ 1 และให้ค่าประสิทธิภาพโดยรวม (F1) เท่ากับ 1

Class S มีข้อมูลรวม 267 เอกสาร ทำนายถูกต้องทั้งหมด 266 เอกสาร และ ทำนายไม่ถูกต้องรวม 1 เอกสาร โดยทำนายเป็น Class A ผลการทดลองสามารถอธิบายคำตอบการค้นคืนเอกสารและให้ค่าความแม่นยำ (Precision) เท่ากับ 1 ให้ค่าระลึก (Recall) เท่ากับ 1 และให้ค่าประสิทธิภาพโดยรวม (F1) เท่ากับ 1

Class A มีข้อมูลรวม 201 เอกสาร ทำนายถูกต้องทั้งหมด ผลการทดลองสามารถอธิบายคำตอบการค้นคืนเอกสารและให้ค่าความแม่นยำ (Precision) เท่ากับ 1 ให้ค่าระลึก (Recall) เท่ากับ 1 และให้ค่าประสิทธิภาพโดยรวม (F1) เท่ากับ 1

Class N มีข้อมูลรวม 245 เอกสาร ทำนายถูกต้องทั้งหมด 242 เอกสาร และ ทำนายไม่ถูกต้องรวม 2 เอกสาร โดยทำนายเป็น Class E จำนวน 2 เอกสาร และ Class S จำนวน 1 เอกสาร ผลการทดลองสามารถอธิบายคำตอบการค้นคืนเอกสารและให้ค่าความแม่นยำ (Precision) เท่ากับ 0.99 ให้ค่าระลึก (Recall) เท่ากับ 0.99 และให้ค่าประสิทธิภาพโดยรวม (F1) เท่ากับ 0.99

จากผลการทดลอง เมื่อพิจารณาผลการทำนายทั้ง 4 Class พบว่ามีการทำนายคลาดเคลื่อนจำนวน 3 Class ประกอบด้วย Class E, S, และ N โดยสามารถสรุปสาเหตุของความคลาดเคลื่อนได้ใน 3 ประเด็นหลัก ดังนี้ 1) เนื่องจากรายการคำถามทั้ง 4 กลุ่มมีความคล้ายกันมากเกินไป 2) เมื่อพิจารณาเปรียบเทียบคำสำคัญที่เป็นลักษณะเด่นระหว่างรายการคำถามและคำตอบมีความใกล้เคียงกันมากแต่ไม่ใช่คำเดียวกัน ยกตัวอย่างเช่น “การท่องเที่ยว (karn thong theiw)”, “เที่ยว (theiw)”, “ท่องเที่ยว (thong theiw)”, “สถานที่ท่องเที่ยว (satharn thi thong theiw)”, และ “แหล่งท่องเที่ยว (hlaeng thong theiw)” จึงส่งผลให้นายทำนายคลาดเคลื่อน และ 3) ผลการค้นคืนจากคำที่เป็นตัวแทนของเอกสารคำตอบนั้นไม่สามารถอธิบายความหมายของคำที่ค้นคืนมาได้ ซึ่งเป็นจุดด้อยของวิธีการแบบ TF-IDF



บทที่ 6

สรุปและอภิปรายผลการวิจัย

6.1 สรุปผลการวิจัย

การวิจัยในครั้งนี้ประกอบด้วย 3 วัตถุประสงค์หลัก ในส่วนของการสรุปผลการวิจัยสามารถอธิบายการสรุปผลการวิจัยตามวัตถุประสงค์ดังนี้

6.1.1 วัตถุประสงค์ข้อแรก

เพื่อวิเคราะห์องค์ประกอบของกลุ่มรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ จากผลการวิจัยสามารถอภิปรายผลการวิเคราะห์กลุ่มรายการคำถามเพื่อนำมาเป็นฐานและแนวทางสำหรับการออกแบบหุ่นยนต์สนทนา โดยได้นำรายการคำถามที่มีค่า Factor loading สูงที่สุดมาจัดเรียงลำดับ 1-3 และกำหนดชื่อกลุ่มใหม่ใน 4 กลุ่ม สามารถอภิปรายผลการวิจัยดังนี้

กลุ่มที่ 1 กลุ่มรายการคำถามด้านการให้บริการที่เป็นเลิศ (Excellent service) ประกอบด้วย 1) Sale & Promotion (ร้านค้ากำหนดราคาสินค้าและบริการได้อย่างเหมาะสม) 2) Product & Service (ร้านค้ามีสินค้าและบริการที่หลากหลาย) และ 3) Members & Parking (แหล่งท่องเที่ยวมีสถานที่จอดรถที่เหมาะสมและปลอดภัย)

กลุ่มที่ 2 กลุ่มรายการคำถามด้านสิ่งอำนวยความสะดวกที่มีมาตรฐาน (Standard facilities) ประกอบด้วย 1) Accommodation (ที่พักมีความปลอดภัย) 2) Transportation (เข้าถึงที่พักได้อย่างสะดวก) และ Booking (ความหลากหลายของประเภทที่พัก) ซึ่งสอดคล้องกับงานวิจัยของ [95] ความต้องการที่แตกต่างและพฤติกรรมที่เปลี่ยนไปของผู้บริโภคด้านสินค้าและบริการ เป็นรากฐานสำคัญที่ส่งผลกระทบต่ออุตสาหกรรมท่องเที่ยวและบริการ รวมทั้งการเสริมสร้างศักยภาพเครือข่ายในเชิงพื้นที่เพื่อให้เกิดความคล่องตัวของธุรกิจและผลประโยชน์ที่เกื้อกูลกันสู่การพัฒนาการท่องเที่ยวอย่างยั่งยืน

กลุ่มที่ 3 กลุ่มรายการคำถามด้านข้อมูลข่าวสารที่ถูกต้องและทันสมัย (Accuracy of information) ประกอบด้วย 1) Tourist info (แหล่งท่องเที่ยวมีระบบสารสนเทศและเว็บไซต์ให้ข้อมูลกับนักท่องเที่ยว) 2) Travel guide (มีคู่มือหรือเอกสารเผยแพร่การท่องเที่ยว และมีรายละเอียดต่าง ๆ ครบถ้วน) 3) Contact us (มีการประชาสัมพันธ์กิจกรรมการท่องเที่ยวผ่าน Social media อย่างต่อเนื่อง) สอดคล้องกับผลการวิจัยของ [96] [95] ในด้านข้อมูลเชิงพื้นที่ที่มีความพร้อมและความสมบูรณ์จะช่วยส่งเสริมการท่องเที่ยว และส่งผลต่อความสะดวกในการเข้าถึงข้อมูลที่รวดเร็วถูกต้อง

รวมถึงการนำมาซึ่งการตัดสินใจเพื่อการท่องเที่ยว และ [97] ด้านความได้เปรียบคู่แข่งขั้นทางธุรกิจ การท่องเที่ยว

กลุ่มที่ 4 กลุ่มรายการคำถามด้านการมีวัฒนธรรมที่โดดเด่นดิงาม (Noble culture) ประกอบด้วย 1) Cultural (ชุมชนมีแหล่งท่องเที่ยวที่เป็นเอกลักษณ์ (Uniqueness) โดดเด่น) 2) Festival & Activities (ชุมชนมีแหล่งท่องเที่ยวที่มีอัตลักษณ์ (Identity) เฉพาะท้องถิ่น) และ 3) History (ชุมชนมีแหล่งท่องเที่ยวทางประวัติศาสตร์ที่สะท้อนวัฒนธรรม วิถีชีวิต และประเพณี) ในงานวิจัย [95] ยังกล่าวถึงศักยภาพด้านวัฒนธรรมที่โดดเด่นนั้นหากคนในพื้นที่ขาดซึ่งความสำนึกในคุณค่าและเห็นแก่ตัวจะนำมาซึ่งการปิดกั้นชุมชนและส่งผลกระทบต่อการท่องเที่ยวโดยชุมชนในระยะยาว

สรุปได้ว่า ปัจจัยที่แสดงถึงความต้องการที่จำเป็นสำหรับการท่องเที่ยวที่ประสบความสำเร็จ ของการท่องเที่ยวโดยชุมชนเขตอีสานใต้ โดยการหาความสัมพันธ์ของโครงสร้างของตัวแปร ได้กลุ่มรายการคำถามที่มีความสัมพันธ์กัน (ESAN model) 4 กลุ่ม ประกอบด้วย ด้านการให้บริการที่เป็นเลิศ (Excellent service) ด้านสิ่งอำนวยความสะดวกที่มีมาตรฐาน (Standard facilities) ด้านข้อมูลข่าวสารที่ถูกต้องและทันสมัย (Accuracy of information) และ ด้านการมีวัฒนธรรมที่โดดเด่นดิงาม (Noble culture) เพื่อนำมาเป็นฐานและแนวทางสำหรับการออกแบบหุ่นยนต์สนทนา ESANBOT

6.1.2 วัตถุประสงค์ข้อที่สอง

เพื่อค้นหารูปแบบคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ การสังเคราะห์ผลจากกฎความสัมพันธ์ แสดงให้เห็นถึงรูปแบบและลำดับความสัมพันธ์การสอบถาม เพื่อแสวงหาข้อมูลสารสนเทศของนักท่องเที่ยวในพื้นที่การท่องเที่ยวโดยชุมชนเขตอีสานใต้ ผู้การนำไปใช้ประโยชน์เชิงวิชาการในประเด็นที่น่าสนใจ คือ นำเสนอรูปแบบการเตรียมข้อมูลเพื่อสารสนเทศที่รวดเร็วและตรงกับความต้องการ ซึ่งสอดคล้องกับทฤษฎีลำดับขั้นความต้องการด้านท่องเที่ยวของมาสโลว์ (Maslow's hierarchy of needs for tourism) [25] [33] โดยสามารถอธิบายรูปแบบรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ [23] [25] [26] [33] [96] บนพื้นฐานของอีสานโมเดล (ESAN model) [22] สรุปสาระสำคัญจากกฎความสัมพันธ์ได้ดังนี้

1) ด้านความปลอดภัย (Safety needs) และ ด้านความต้องการพื้นฐานร่างกาย (Physiological needs) สามารถอธิบายรูปแบบและแรงจูงใจที่มีอิทธิพลต่อการตัดสินใจเดินทางของนักท่องเที่ยว ใน 2 กลุ่ม คือ

กลุ่มที่ 1 การบริการที่เป็นเลิศ (E: Excellent service) โดยนักท่องเที่ยวสนใจสอบถามในประเด็นความปลอดภัย (Security) โดยเริ่มคำถามแรกคือ ความสะอาดและห้องน้ำที่ได้มาตรฐาน (Clean & Standardized toilets) สินค้าและบริการที่หลากหลาย (Products & Services variously) สินค้าและบริการที่เพียงพอ (Products & Services enough) ราคาของสินค้า (Prices of products) ร้านอาหารที่ถูกสุขอนามัย (Hygiene standards restaurant) อาหารพื้นถิ่น (Local food) และ

กลุ่มที่ 2 สิ่งอำนวยความสะดวกที่มีมาตรฐาน (Standard facilities: S) ประเด็นคำถามจะให้ความสำคัญไปที่ความปลอดภัย (Security) โดยเริ่มคำถามแรกคือ ความสะอาดและห้องน้ำที่ได้มาตรฐาน (Clean & Standardized toilets) สินค้าและบริการที่หลากหลาย (Products & Services variously) สินค้าและบริการที่เพียงพอ (Products & Services enough) ราคาของสินค้า (Prices of products) ร้านอาหารที่ถูกสุขอนามัย (Hygiene standards restaurant) อาหารพื้นถิ่น (Local food)

2) ความต้องการความรักและการเป็นส่วนหนึ่งของสังคม (Love and belonging needs) และความต้องการได้รับการยกย่อง หรือนับถือจากบุคคลอื่น (The esteem needs) สอดคล้องกับแรงจูงใจต่อนักท่องเที่ยวและสามารถอธิบายในกลุ่มข้อมูลสารสนเทศที่ถูกต้องและตรงกับความต้องการ (Accuracy of information: A) ซึ่งสะท้อนให้เห็นจากกฎความสัมพันธ์ที่นักท่องเที่ยวให้ความสนใจถามคือ เจ้าหน้าที่ให้คำแนะนำและบริการที่ดีและมีคุณภาพ (Suggestions & Service mind) เริ่มต้นคำถามจากระเบียบ กฎและกติกา (Rules & Regulations) ช่องทางการสื่อสารเช่น เว็บไซต์ (Websites) สื่อสังคมออนไลน์ (Social media) คู่มือการท่องเที่ยว (Tourist guide)

3) ด้านความต้องการความสำเร็จและการเข้าใจตัวเองอย่างแท้จริง (Self-actualization) โดยสามารถอธิบายพฤติกรรมและแรงจูงใจที่มีอิทธิพลต่อการตัดสินใจเดินทางของนักท่องเที่ยวในกลุ่มประเพณีและวัฒนธรรมที่โดดเด่น (Noble culture: N) ซึ่งจะเห็นได้จากกฎความสัมพันธ์ที่แสดงให้เห็นจากการสอบถามที่พบบ่อยของนักท่องเที่ยวจะให้ความสำคัญในประเด็น แหล่งท่องเที่ยวทางประวัติศาสตร์ที่สะท้อนวัฒนธรรม วิถีชีวิต และประเพณี ของชุมชน โดยคำถามของนักท่องเที่ยวจะเริ่มจาก การเข้าถึงแหล่งท่องเที่ยว (Tourist attractions) อัตลักษณ์ (Identity) เอกลักษณ์ (Uniqueness) และการส่งเสริมและอนุรักษ์วัฒนธรรมท้องถิ่น (Conserves the local cultures)

6.1.3 วัตถุประสงค์ข้อที่สาม

เพื่อพัฒนากรอบการทำงานการหาคำตอบสำหรับตอบคำถามอัตโนมัติการท่องเที่ยวโดยชุมชนอีสานได้นั้น การเข้าถึงข้อมูลของนักท่องเที่ยวและผู้มีส่วนเกี่ยวข้องได้อย่างมีประสิทธิภาพ

เป็นความท้าทายของงานวิจัยในครั้งนี้ วิธีการหาค่าความคล้ายคลึงด้วยเทคนิค Cosine similarity บนพื้นฐาน VSM [85] [94] เป็นที่นิยมและมีประสิทธิภาพ โดยนำจุดเด่นของ Text vectorize มาคำนวณให้ได้คุณลักษณะสำคัญเพื่อเป็นตัวแทนเอกสารที่มีคุณภาพ [87] [89] [91] ผลลัพธ์ที่ได้รับในขั้นตอนนี้คือ คลังคำศัพท์ 19,501 คำ จากจำนวนทั้งสิ้น 1,237 เอกสาร ผลการทดลองพบว่า การค้นคืนเอกสารด้วยวิธีการ Cosine similarity และการประเมินประสิทธิภาพของชุดข้อมูลคำตอบทั้งหมด จาก 4 กลุ่ม (Class) ให้ค่าความถูกต้อง (Accuracy) ร้อยละ 99 ประกอบด้วย Class E Class S Class A และ Class N โดยสามารถอธิบายคำตอบจากการค้นคืนเอกสารหรือประสิทธิภาพได้ดีที่สุดตามลำดับดังนี้

Class A โดยให้ค่าความแม่นยำเท่ากับ 1 ให้ค่าระลอกเท่ากับ 1 และให้ค่าประสิทธิภาพโดยรวม เท่ากับ 1

Class S ให้ค่าความแม่นยำเท่ากับ 1 ให้ค่าระลอกเท่ากับ 1 และให้ค่าประสิทธิภาพโดยรวม เท่ากับ 1

Class E ให้ค่าความแม่นยำเท่ากับ 1 ให้ค่าระลอกเท่ากับ 1 และให้ค่าประสิทธิภาพโดยรวม เท่ากับ 1 และ

Class N โดยให้ค่าความแม่นยำเท่ากับ 0.99 ให้ค่าระลอกเท่ากับ 0.99 และให้ค่าประสิทธิภาพโดยรวมเท่ากับ 0.99 ตามลำดับ

6.2 อภิปรายผลการวิจัย

การวิเคราะห์องค์ประกอบของกลุ่มรายการคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานได้ ผลการวิจัยพบว่า องค์ประกอบของกลุ่มรายการคำถาม มี 4 กลุ่มหลัก คือ กลุ่มรายการคำถามด้านการให้บริการที่เป็นเลิศ (Excellent service) 2) กลุ่มรายการคำถามด้านสิ่งอำนวยความสะดวกที่มีมาตรฐาน (Standard facilities) 3) กลุ่มรายการคำถามด้านข้อมูลสารสนเทศและประชาสัมพันธ์ (Accuracy of information) และ 4) กลุ่มรายการคำถามด้านอัตลักษณ์ ประเพณีและวัฒนธรรม (Noble culture) เนื่องจากเป็นองค์ประกอบและปัจจัยสำคัญของการท่องเที่ยว (5's component of tourism) [25] และการท่องเที่ยวโดยชุมชน [7] [8] ที่แสดงให้เห็นถึงความต้องการที่จำเป็นและส่งผลกระทบต่อความสำเร็จของอุตสาหกรรมการท่องเที่ยวในภาพรวม ซึ่งสอดคล้องกับแนวคิดแรงจูงใจทางการท่องเที่ยว (Tourist motivation) [27] Šimková และคณะ [27] โดยการประยุกต์ใช้แนวคิด 5 ลำดับขั้นความต้องการของ Maslow (The 5 levels of Maslow's hierarchy of needs)

การค้นหารูปแบบคำถามที่มีความสัมพันธ์ต่อการท่องเที่ยวโดยชุมชนอีสานใต้ ผลการวิเคราะห์หากฎความสัมพันธ์ (Association rule mining: ARM) ด้วยอัลกอริทึม Apriori โดยพิจารณาเลือกกฎความสัมพันธ์ที่ดีที่สุด (Best rules) จากตัวชี้วัดประสิทธิภาพจากค่า Confidence ที่ยอมรับได้ตั้งแต่ 98% ขึ้นไป และพิจารณาเลือกกฎที่มีลักษณะเด่น มีความเฉพาะที่เป็นไปในทิศทางเดียวกันในเชิงประจักษ์และสอดคล้องกับแนวคิด 5 ลำดับชั้นความต้องการของ Maslow [27] โดยคัดกฎที่มีความซ้ำซ้อนทิ้ง [40] ผลการทดลองพบว่า สามารถสร้างกฎที่ดีที่สุด [99] [100] โดยเรียงตามลำดับความสำคัญจากกฎความสัมพันธ์ของทั้ง 4 กลุ่ม ประกอบด้วย กลุ่มที่ 1 รายการคำถามด้านการให้บริการที่เป็นเลิศ (Excellent service) ให้ความสำคัญใน 3 ประเด็น คือ ความปลอดภัย (Security) อาหาร (Food) ร้านอาหาร (Restaurant) และ สินค้าและบริการ (Products & Services) กลุ่มที่ 2 รายการคำถามด้านสิ่งอำนวยความสะดวกที่มีมาตรฐาน (Standard facilities) ให้ความสำคัญใน 2 ประเด็น คือ ความปลอดภัย (Security) และ กิจกรรมสำหรับการท่องเที่ยว (Activities) กลุ่มที่ 3 รายการคำถามด้านข้อมูลสารสนเทศและประชาสัมพันธ์ (Accuracy of information) ให้ความสำคัญใน 3 ประเด็น คือ การให้คำแนะนำและบริการ (Suggestions & Service) และ ช่องทางการสื่อสารประชาสัมพันธ์ (Channels to communicate and social media) และ กลุ่มที่ 4 รายการคำถามด้านอัตลักษณ์ ประเพณีและวัฒนธรรม (Noble culture) ให้ความสำคัญใน 3 ประเด็น คือ ประวัติศาสตร์ (Historical) การเข้าถึงแหล่งท่องเที่ยว (Tourist attractions) อัตลักษณ์ของชุมชน (Identity) และการอนุรักษ์ไว้ซึ่งประเพณีท้องถิ่น โดยการนำรูปแบบรายการคำถามที่ได้เป็นแนวทางสำหรับการเตรียมข้อมูลล่วงหน้าเพื่อเป็นคำตอบ ให้ตรงกับความต้องการของนักท่องเที่ยว [40] ซึ่งเป็นการประยุกต์ใช้กับแนวคิดการวิเคราะห์ตะกร้าสินค้า (Market basket analysis) [101] และยังเป็นที่ยอมรับในปัจจุบัน

การพัฒนากรอบการทำงานการหาคำตอบสำหรับตอบคำถามอัตโนมัติการท่องเที่ยวโดยชุมชนอีสานใต้ ผลการวิจัยพบว่า โมเดลที่พัฒนาขึ้นสามารถสกัดคุณลักษณะสำคัญที่ใช้เป็นตัวแทนเอกสารได้อย่างมีประสิทธิภาพ [87] [89] [91] โดยได้คลังคำศัพท์ 17,180 คำ จากจำนวนทั้งสิ้น 1,237 เอกสาร การค้นคืนเอกสารด้วยวิธีการ Cosine similarity และการประเมินประสิทธิภาพของชุดข้อมูลคำตอบทั้งหมด จาก 4 กลุ่ม (Class) ให้ค่าความถูกต้อง (Accuracy) ร้อยละ 97 โดยที่ Class S ให้ค่าความแม่นยำ (Precision) เท่ากับ 0.97 ให้ค่าระลึก (Recall) เท่ากับ 1 และให้ค่าประสิทธิภาพโดยรวม (F1) เท่ากับ 0.98 ซึ่งเป็นสามารถอธิบายคำตอบจากการค้นคืนเอกสารหรือประสิทธิภาพการค้นคืนได้ดีที่สุด

การออกแบบรายการคำถามหรือการเตรียมคำถามที่ดี นำสู่การได้มาซึ่งคำตอบและการเตรียมข้อมูลคำตอบได้ล่วงหน้า ซึ่งเป็นองค์ประกอบสำคัญของการออกแบบระบบหุ่นยนต์สนทนาอัตโนมัติ [36] [52] [53] [54] ให้เข้าใจภาษาธรรมชาติและความต้องการที่แท้จริงในบทสนทนาของนักท่องเที่ยว เป็นพื้นฐานที่สำคัญให้ Chatbot ได้เรียนรู้สำหรับการตอบคำถามได้อย่างถูกต้อง [55] ตรงประเด็น กระชับ ง่ายต่อความเข้าใจและสื่อสารได้อย่างรวดเร็ว ส่งผลต่อการตัดสินใจเดินทางทางท่องเที่ยวและเป็นฐานในการออกแบบและพัฒนาหุ่นยนต์สนทนา (Chatbot) ให้สามารถสื่อสารกับมนุษย์ผ่านช่องทางออนไลน์แบบเรียลไทม์ (Real-time) [56] ซึ่งมีวัตถุประสงค์คล้ายกันในการศึกษาวิธีการออกแบบและพัฒนาระบบถามตอบอัตโนมัติ สอดคล้องกับงานวิจัยของ Phuc Do และคณะ [102] ในปี 2021 ได้พัฒนาระบบการตอบคำถามการท่องเที่ยวของเวียดนามด้วย Knowledge graph and deep learning เพื่อตอบคำถามของนักท่องเที่ยว เช่น ชื่อแหล่งท่องเที่ยว ความเป็นมาของการตั้งชื่อ พร้อมระบบแนะนำการท่องเที่ยว เช่นเดียวกับงานวิจัยของ Jiahui Li และคณะ [103] ในปี 2022 ได้พัฒนารอบการทำงานคำถามคำตอบอัตโนมัติบนหลักการ Knowledge Graph จากเว็บไซต์ด้านการท่องเที่ยว จากผลการทดลองพบว่า ระบบถามตอบอัตโนมัติสามารถตอบคำถามได้เป็นอย่างดี มีการรู้จำถึงเจตนาของการค้นคืน รวมไปถึงการแสดงผลของคำตอบที่ได้เป็นอย่างดีมีประสิทธิภาพ และในปีเดียวกัน Myeong-Cheol Jwa และคณะ [104] นำเทคนิค Named Entity Recognition เพื่อวิเคราะห์และจำแนกหน่วยคำภาษาเกาหลีและอังกฤษ เทรนด้วยโมเดล KoBERT-CRF NER model เพื่อพัฒนาแอปพลิเคชันด้านการท่องเที่ยว ผลการทดลองพบว่าได้ค่า F1 ที่ 0.94 จากการทบทวนวรรณกรรมที่เกี่ยวข้องสะท้อนให้เห็นถึงจุดเด่นของงานวิจัยในครั้งนี้ ในการประยุกต์ใช้กับแนวคิดการวิเคราะห์ตะกร้าสินค้า (Market basket analysis) [101] เพื่อให้ได้มาซึ่งรูปแบบรายการคำถามเพื่อใช้เป็นแนวทางสำหรับการเตรียมข้อมูลล่วงหน้าเพื่อเป็นคำตอบ ให้ตรงกับความต้องการของนักท่องเที่ยว [40] ให้การบริการข้อมูลข่าวสารอย่างครอบคลุมและเข้าถึงกลุ่มนักท่องเที่ยวได้ตลอดเวลาอย่างมีประสิทธิภาพ เพื่อสร้างความมั่นใจและโอกาสต่อการตัดสินใจเดินทางของนักท่องเที่ยวในลำดับต่อไป

บรรณานุกรม

- [1] P. Buppatanakorn, *Community-Base tourism*, 2nd., vol. 2nd. Bangkok: Cocoon and Co. Limited, 2017.
- [2] United Nations, “UN sustainable development goals,” *United Nations*, 2021.
<https://mineraodados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf> (accessed Jan. 26, 2021).
- [3] O. S. Kolbasov, “UN Conference on Environment and Development,” *Izv. - Akad. Nauk. Seriya Geogr.*, vol. 6, no. June, pp. 47–54, 1992, doi: 10.4135/9781412971867.n128.
- [4] M. KostiĆ and M. JovanoviĆ TonČev, “Importance of sustainable tourism,” no. February, pp. 722–725, 2014, doi: 10.15308/sinteza-2014-722-725.
- [5] The National Economic and Social Development Plan, “Transformation to Hi-Value and Sustainable Thailand, The Thirteen Plan (2023-2027),” 2021. doi: 10.11164/jjsps.13.6_956_4.
- [6] P. Jaroenwanit, “Tourism Market Development in the Northeast: A case study of Southern I-SAN,” *KKU Res. J.*, vol. 1, no. 8, pp. 101–110, 2003.
- [7] G. Del Chiappa, M. Atzeni, and V. Ghasemi, “Community-based collaborative tourism planning in islands: A cluster analysis in the context of Costa Smeralda,” *J. Destin. Mark. Manag.*, vol. 8, no. November 2016, pp. 41–48, 2018, doi: 10.1016/j.jdmm.2016.10.005.
- [8] Tourism Authority of Thailand: TAT, “Community-Base Tourism: CBT,” 2021.
- [9] A. Wesoho and S. Kiatwanit, “Information Retrieval on Internet of the Faculty of Liberal Arts’ Students, Rajamangala University of Technology Phra Nakhon,” Rajamangala University of Technology Phra Nakhon, Bangkok, 2012.
- [10] A. Kokkhangplu, “Factors influencing the social inequality of community-based tourism , Thailand.,” *J. Hum. Sci. Fac. Humanit. Chiang Mai Univ.*, vol. 3, no. 21, pp. 214–236, 2020.
- [11] S. Threeranoot, “Tourism Product-Related Purchasing Behaviors and Online Media Marketing Mix of Generation X and Y Tourists.,” *SSRU Grad. Stud. J.*, no. Aug, pp. 1–13, 2020.

- [12] S. Sregongsang, "The development of tourism routes linked to tourism identities for the ways of lower Northeastern part of Thailand.," *J. Lib. Arts, Ubon Ratchathani Univ.*, vol. 14, no. 1, pp. 198–231, 2018, [Online]. Available: https://so03.tci-thaijo.org/index.php/jla_ubu/article/view/243254.
- [13] The World Tourism Organization (UNWTO), "International Tourist Arrivals Reach 1.4 billion Two Years Ahead of Forecasts | UNWTO," *The World Tourism Organization (UNWTO)*, 2019. <https://www.unwto.org/global/press-release/2019-01-21/international-tourist-arrivals-reach-14-billion-two-years-ahead-forecasts> (accessed Oct. 27, 2020).
- [14] P. Onya, "Community Based Tourism Management by Tourist Destination Development Process," *Journal of Humanit. Soc. Sci. Thonburi Univ.*, pp. 45–54, 2020.
- [15] K. Laoyang, "Guidelines for the Development of Competitiveness for Foreign Backpacker Tourists in Suratthani.," *J. Manag. Sci. Chiangrai Rajabhat Univ.*, vol. 1, no. 13, pp. 154–172, 2018.
- [16] B. Hamoud and E. Atwell, "Quran question and answer corpus for data mining with WEKA," pp. 211–216, 2016.
- [17] S. S. Laddha and P. Jawandhiya M., "Onto-Semantic Indian Tourism Information Retrieval System.," *Recent Stud. Comput. Intell.*, no. DoSCI, 2020, doi: 10.1007/978-981-15-8469-5.
- [18] S. S. Ghosh, *Human-centric Faceted Approach for Ontology Construction*, vol. 47, no. 1. KNOWLEDGE ORGANIZATION, 2019.
- [19] N. Jongrak, "Studying the use of travel agencies by Thai residents of the Bangkok metropolitan area.," *Thammasat Univ.*, 2017.
- [20] U. Chaisoong and S. Tirakoat, "The Clustering of Questions Affect to Tourist's Decision Making for Chatbot Design," *17th Int. Conf. Electr. Eng. Comput. Telecommun. Inf. Technol. ECTI-CON 2020*, pp. 784–787, 2020, doi: 10.1109/ECTI-CON49241.2020.9158069.
- [21] U. Chaisoong and S. Tirakoat, "Tourist information-seeking behaviours using association rule mining," *ICIC Express Lett.*, vol. 15, no. 9, pp. 915–923, 2021, doi: 10.24507/icicel.15.09.915.
- [22] UNWTO, *Recommendations on Sustainable Development of Indigenous Tourism*. World Tourism Organization (UNWTO), 2019.

- [23] D. Sharron, "Tourism : an introductory text / Sharron Dickman - Trove," *Hodder Education*, 1996. <https://trove.nla.gov.au/work/5433050> (accessed Oct. 27, 2020).
- [24] R. C. Lucas, "Clawson and Knetsch, Economics of Outdoor Recreation," 1968. Accessed: May 28, 2021. [Online]. Available: <https://digitalrepository.unm.edu/nrj/vol8/iss4/13>.
- [25] E. Šimková and J. Holzner, "ScienceDirect motivation of tourism participants," *Procedia-Social Behav. Sci.*, vol. 159, pp. 660–664, 2014, doi: 10.1016/j.sbspro.2014.12.455.
- [26] J. S. Susan Horner, "Consumer behaviour in tourism.," *Oxford, Butterworth-Heinemann*, 2007. <https://www.economy.gov.ae/Publications/Consumer Behaviour in Tourism.pdf> (accessed Nov. 12, 2020).
- [27] G. D. Israel, "Determining Sample Size," *Food Agric. Sci. IFAS*, pp. 1–5, 1992.
- [28] R. Likert, *A Technique for the Measurement of Attitudes*. New York: New York University, 1932.
- [29] T. Chirachoenwong, "A Study of Factors Influencing Travel Behaviors in Chiang Mai City," in *ICRU International Conference: Sustainable Community Development*, 2019, vol. 1st, p. 16.
- [30] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951, doi: 10.1007/bf02310555.
- [31] J. De Vos, S. Tim, V. Van Acker, and F. Witlox, "How satisfying is the Scale for Travel Satisfaction ?," *Transp. Res. Part F Psychol. Behav.*, vol. 29, pp. 121–130, 2015, doi: 10.1016/j.trf.2015.01.007.
- [32] I.-G. Czibula, "Enhancing relational association rules with gradualness.," *Int. J. Inov. Comput. Inf. Control.*, vol. 13, no. 1, 2017.
- [33] A. M. M. Robert Christie. Mill, "The tourism system : an introductory text," *Prentice-Hall, Inc.*, 1998. <https://archive.org/details/tourismssystemint0000mill/page/n5/mode/2up> (accessed Nov. 17, 2020).
- [34] B. R. Ranoliya, N. Raghuwanshi, and S. Singh, "Chatbot for University related FAQs," pp. 1525–1530, 2017.
- [35] Economics Tourism and Sports Division, "National tourism development plan 2016-2022," 2016. [Online]. Available: <https://www.mots.go.th/>.

- [36] D. Ellis, “A behavioural model for information retrieval system design,” *J. Inf. Sci.*, vol. 15, no. 4–5, pp. 237–247, Aug. 1989, doi: 10.1177/016555158901500406.
- [37] B. D. Wijanarko, Y. Heryadi, H. Toba, and W. Budiharto, “Automated question generating method based on derived keyphrase structures from bloom’s taxonomy,” *ICIC Express Lett.*, vol. 14, no. 11, pp. 1059–1067, 2020, doi: 10.24507/icicel.14.11.1059.
- [38] T. Arreeras, M. Endo, and H. Takahashi, “An Association Rule Mining-Based Exploration of Travel Patterns in Wide Tourism Areas using A Wi-Fi Package Sensing Survey,” vol. 13, pp. 1099–1113, 2019.
- [39] H. Jiawei, *Data Mining Concepts and Techniques.*, 2nd. Cambridge: Morgan Kaufmann Publishers, 2006.
- [40] O. Kolomiyets and M. F. Moens, “A survey on question answering technology from an information retrieval perspective,” *Inf. Sci. (Ny)*, vol. 181, no. 24, pp. 5412–5434, Dec. 2011, doi: 10.1016/j.ins.2011.07.047.
- [41] P. Bangcharoansub and C. Jaikaeo, “A Machine-Translation based Approach to Word Boundary Identification: A Projective Analogy of Bilingual Translation,” 2009.
- [42] T. Horsuwan, K. Kanwatchara, P. Vateekul(B), and B. Kijisirikul, “A Comparative Study of Pretrained Language Models on Thai Social Text Categorization,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, vol. 12033 LNAI, pp. 63–75, doi: 10.1007/978-3-030-41964-6.
- [43] L. Lowphansirikul, C. Polpanumas, N. Jantrakulchai, S. Nutanong, ‡ Pythainlp, and C. P. Org, “WangchanBERTa: Pretraining transformer-based Thai language models,” *Comput. Lang. Cornell Univ.*, 2021, doi: 10.48550/arXiv.2101.09635 Focus to learn more.
- [44] H. Chen, Z. Dai, J. Tang, and M. Li, “Web Evaluation Analysis of Tourism Destinations Based on Data Mining,” *2018 IEEE 4th Int. Conf. Comput. Commun.*, pp. 1803–1808, 2019, doi: 10.1109/compcomm.2018.8781024.
- [45] M. Membiela-Pollan, E. Sanchez-Amboaze, C. Rodriguez-Vazquez, and A. P. Ludena-Reyes, “Social capital, an ‘asset’ for community tourism in Ecuador,” *Iber. Conf. Inf. Syst. Technol. Cist.*, vol. 2019-June, no. June, pp. 19–22, 2019, doi: 10.23919/CISTI.2019.8760779.

- [46] J. Yuan and Y. Liu, "Research on tourism development of cultural and creative industry park based on the network text analysis : AA case study of dafen oil painting village in Shenzhen," *2017 3rd Int. Conf. Inf. Manag. ICIM 2017*, pp. 37–41, 2017, doi: 10.1109/INFOMAN.2017.7950343.
- [47] M. Fazzolari and M. Petrocchi, "A study on online travel reviews through intelligent data analysis," *Inf. Technol. Tour.*, vol. 20, no. 1–4, pp. 37–58, 2018, doi: 10.1007/s40558-018-0121-z.
- [48] L. Li, K. Y. Lee, and S. B. Yang, "Exploring the effect of heuristic factors on the popularity of user-curated 'Best places to visit' recommendations in an online travel community," *Inf. Process. Manag.*, vol. 56, no. 4, pp. 1391–1408, 2019, doi: 10.1016/j.ipm.2018.03.009.
- [49] T. Ministry of Tourism and Sports, "Action Plan for Tourism Development within the South Northeastern Civilization Tourism Development Zone (2016-2020)," 2558.
- [50] A. Rajagopal and N. Vedamanickam, "New Approach to Human AI Interaction to Address Digital Divide AI divide: Creating an Interactive AIplatform to Connect Teachers Students," *Proc. 2019 3rd IEEE Int. Conf. Electr. Comput. Commun. Technol. ICECCT 2019*, pp. 1–6, 2019, doi: 10.1109/ICECCT.2019.8869174.
- [51] V. Wudaru, N. Koditala, A. Reddy, and R. Mamidi, "Question answering on structured data using NLIDB approach," *2019 5th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2019*, pp. 1–4, 2019, doi: 10.1109/ICACCS.2019.8728487.
- [52] R. Kincaid and G. Pollock, "Nicky : Toward A Virtual Assistant for Test and Measurement Instrument Recommendations," *2017 IEEE 11th Int. Conf. Semant. Comput.*, pp. 196–203, 2017, doi: 10.1109/ICSC.2017.11.
- [53] Y. Li, H. Li, J. Li, N. Zhang, and G. Yuan, "A Deep Semantic Matching Network for Answer Selection," *2019 Chinese Control Conf.*, no. 1, pp. 8683–8688, 2019.
- [54] D. Arteaga, J. Arenas, F. Paz, M. Tupia, and M. Bruzza, "Design of information system architecture for the recommendation of tourist sites in the city of Manta, Ecuador through a Chatbot," *Iber. Conf. Inf. Syst. Technol. Cist.*, vol. 2019-June, no. June, pp. 19–22, 2019, doi: 10.23919/CISTI.2019.8760669.

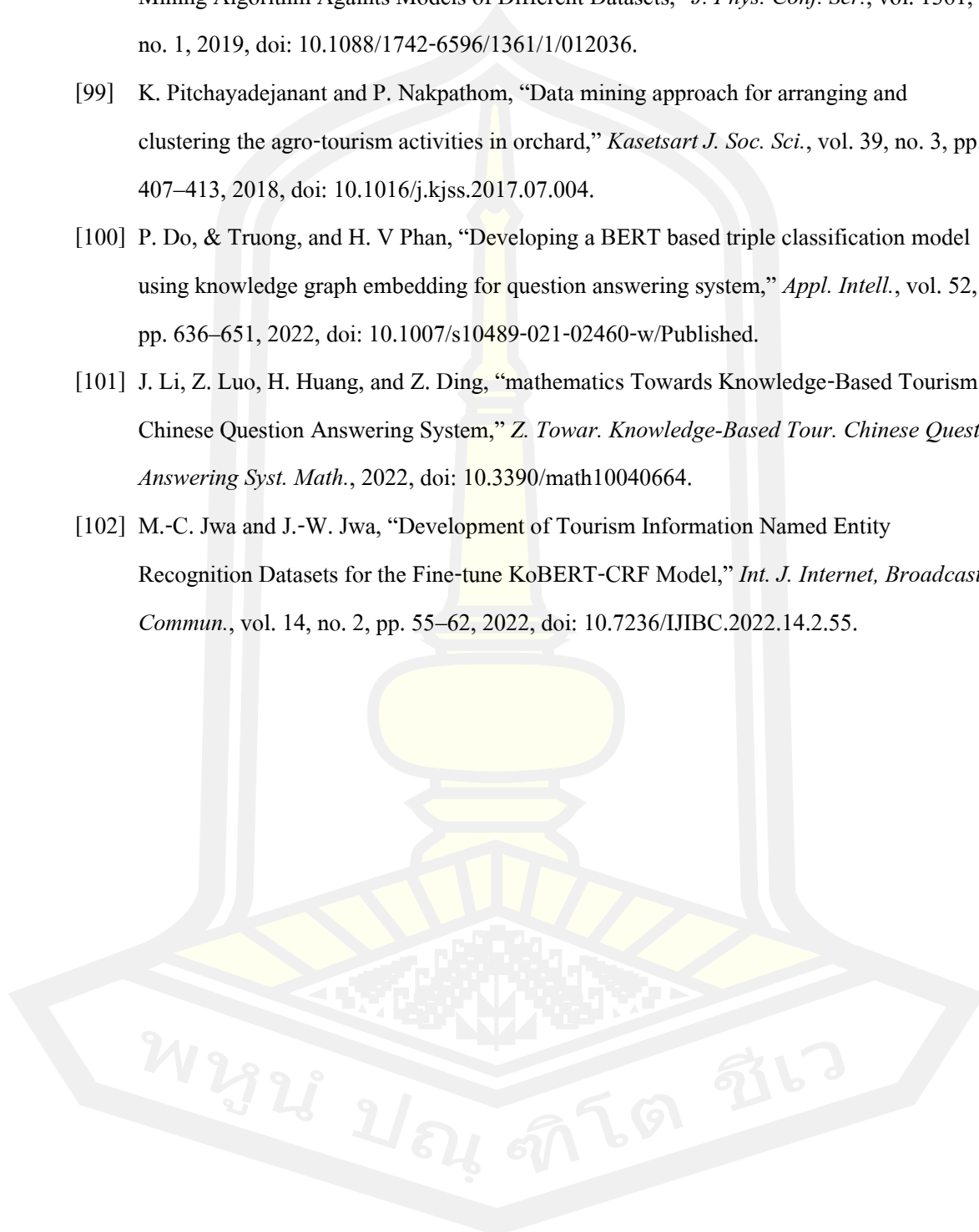
- [55] R. W. Butler, "Sustainable tourism: A state-of-the-art review," *Tour. Geogr.*, vol. 1, no. 1, pp. 7–25, 1999, doi: 10.1080/14616689908721291.
- [56] N. Phopi, "The problem and obstacle of market expansion in less visited area for Russian tourist: A case study of Phitsanulok province.," *Faculty of Liberal Arts, Thammasat University*, 2018. <http://km-ir.arts.tu.ac.th/s/ir/item/1216> (accessed Jan. 12, 2021).
- [57] V. T. C. Middleton, "Marketing in travel and tourism.," *Butterworth Heineman*, 1994. https://books.google.co.th/books/about/Marketing_in_Travel_and_Tourism.html?id=RVoXAQAAMAAJ&redir_esc=y (accessed Oct. 27, 2020).
- [58] E. R. Malinowski, "Theory of Error in Factor Analysis," *Anal. Chem.*, vol. 49, no. 4, pp. 606–612, 1977, doi: 10.1021/ac50012a026.
- [59] B. D. Haig, "Exploratory factor analysis, theory generation, and scientific method," *Stud. Appl. Philos. Epistemol. Ration. Ethics*, vol. 45, no. August 2013, pp. 65–88, 2018, doi: 10.1007/978-3-030-01051-5_4.
- [60] H. Taherdoost, S. Sahibuddin, and N. Jalaliyoon, "Exploratory Factor Analysis ; Concepts and Theory To cite this version : HAL Id : hal-02557344 Exploratory Factor Analysis ; Concepts and Theory," *Math. Comput. Sci. Eng. Ser.*, no. 27, pp. 375–382, 2014.
- [61] M. Scholz, "Approaches to analyse and interpret biological profile data.," Max Planck Institute of Molecular Plant Physiology, 2006.
- [62] A. Prutzkow, "Algorithms and data structures for Association rule mining and its complexity analysis," Nov. 2018, pp. 558–568, doi: 10.15405/epsbs.2018.11.02.62.
- [63] A. Aninditya, M. A. Hasibuan, and E. Sutoyo, *Text Mining Approach Using TF-IDF and Naive Bayes for Classification of Exam Questions Based on Cognitive Level of Bloom's Taxonomy.* .
- [64] P. Songram, *Pattern mining: Theory and practice*, 1st ed. Mahasarakham: Rajaphat Mahasarakham Univ., 2020.
- [65] C. Jareanpon, "Multi-disciplinary Trends in Artificial Intelligence," in *Book Cover and Content Similarity Retrieval Using Computer Vision and NLP Techniques.*, 2021, pp. 34–44, doi: 10.1007/978-3-030-80253-0_4.
- [66] C. D. Manning, *An Introduction to Information Retrieval.*, Online edi., no. Online edition (c) 2009 Cambridge U. Cambridge University, 2009.

- [67] H. S. Manning, Christopher D., Prabhakar Raghavan., *An Introduction to Information Retrieval.*, Online edi., no. Online edition. Cambridge: Cambridge University Press, 2009.
- [68] K. Jairak, “Applying Ontology based Recommendation System for Community Ecotourism: A Case Study of Mae Kam Pong Village, Mae On Sub-District, Chiang Mai Province,” *J. Mod. Manag. Sci.*, vol. 9, no. 2, pp. 146–160, 2016.
- [69] N. Chayangkoon, “An Applied Data Mining Technique to Develop an Information Storage and Retrieval for Intelligence of Locality for the Southern Part of Thailand in Suratthani Province.,” *J. KMUTNB*, vol. 24, no. 2, pp. 374–384, 2014.
- [70] P. Titiporn, Chansiriwat Achainee, “The Development of Ontology for Health Tourism in Thailand.” *FEU Academic Review*, pp. 113–129, 2018.
- [71] K. She, “Tourism Recommender System Utilising Property Graph Ontology as Knowledge Base,” *Assoc. Comput. Mach. (ACM).*, vol. ICCMS '20, no. Feb 26-28, pp. 14–18, 2020, doi: 10.1145/3408066.3408102.
- [72] I. Chuenta, J. Thongkam, and J. Phuboon-ob, “Information Retrieval In Thai Northeast Travel Utilizing Ontology.,” *Inf. Technol. J.*, vol. 10, no. 2, pp. 15–25, 2014.
- [73] C. Haruechaiyasak, S. Kongyoung, and M. N. Dailey, “A Comparative Study on Thai Word Segmentation Approaches,” 2008, doi: 10.1109/ECTICON.2008.4600388.
- [74] W. Phatthiyaphaibun, K. Chaovavanich, C. Polpanumas, A. Suriyawongkul, L. Lowphansirikul, and A. Chormai, “PyThaiNLP: Thai Natural Language Processing in Python.,” *PyThaiNLP documentation*, 2016. <https://pythainlp.github.io/docs/2.3/> (accessed Mar. 18, 2022).
- [75] J. Intasorn, S. Gertphol, and U. Sammapun, “Thai Sentiment Lexicon Construction,” in *2021 13th International Conference on Knowledge and Smart Technology (KST)*, Jan. 2021, pp. 123–128, doi: 10.1109/KST51265.2021.9415804.
- [76] P. Saipech and P. Seresangtakul, “Automatic Thai Subjective Examination using Cosine Similarity,” *ICAICTA 2018 - 5th Int. Conf. Adv. Informatics Concepts Theory Appl.*, pp. 214–218, 2018, doi: 10.1109/ICAICTA.2018.8541276.
- [77] L. Augusto and S. Pizzato, “Using Linguistically Motivated Features in Document Retrieval for Question Answering,” no. July, 2009.

- [78] K. Kesorn, *Information retrieval system: Concepts and future directions.*, 1st ed. Phitsanulok: Department of Computer Science and Information Technology in Science Faculty, Naresuan University, 2015.
- [79] C. Sirichanya and K. Kraisak, "Semantic data mining in the information age : A systematic review," *Int. J. Intell. Syst.*, vol. 36, no. 8, pp. 3880–3916, 2021, doi: 10.1002/int.22443.
- [80] J. Wang and Y. Dong, "Measurement of text similarity: A survey," *Inf.*, vol. 11, no. 9, pp. 1–17, 2020, doi: 10.3390/info11090421.
- [81] R. Singh and S. Singh, "Text Similarity Measures in News Articles by Vector Space Model Using NLP," *J. Inst. Eng. Ser. B*, vol. 102, no. 2, pp. 329–338, Apr. 2021, doi: 10.1007/s40031-020-00501-5.
- [82] M. Eklund, "Comparing Feature Extraction Methods and Effects of Pre-Processing Methods for Multi-Label Classification of Textual Data," *Degree Proj. Comput. Sci. Eng.*, 2018.
- [83] R. K. Mishra, S. Urolagin, and A. A. Jothi J, "A Sentiment analysis-based hotel recommendation using TF-IDF Approach," in *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, Dec. 2019, pp. 811–815, doi: 10.1109/ICCIKE47802.2019.9004385.
- [84] A. D. Afrizal, N. A. Rakhmawati, and A. Tjahyanto, "New Filtering Scheme Based on Term Weighting to Improve Object Based Opinion Mining on Tourism Product Reviews," *Procedia Comput. Sci.*, vol. 161, pp. 805–812, 2019, doi: 10.1016/j.procs.2019.11.186.
- [85] W. Chen, Z. Xu, X. Zheng, Q. Yu, and Y. Luo, "Research on Sentiment Classification of Online Travel Review Text," *Appl. Sci.*, vol. 10, no. 15, p. 5275, Jul. 2020, doi: 10.3390/app10155275.
- [86] N. Khamphakdee and P. Seresangtakul, "A Framework for Constructing Thai Sentiment Corpus using the Cosine Similarity Technique," *KST 2021 - 2021 13th Int. Conf. Knowl. Smart Technol.*, pp. 202–207, 2021, doi: 10.1109/KST51265.2021.9415802.
- [87] M. Eminagaoglu, "A new similarity measure for vector space models in text classification and information retrieval," *J. Inf. Sci.*, 2020, doi: 10.1177/0165551520968055.

- [88] P. Meesad, "Thai Fake News Detection Based on Information Retrieval, Natural Language Processing and Machine Learning," vol. 2, p. 425, 2021, doi: 10.1007/s42979-021-00775-6.
- [89] X. Yang, K. Yang, T. Cui, M. Chen, and L. He, "A Study of Text Vectorization Method Combining Topic Model and Transfer Learning," 2022, doi: 10.3390/pr10020350.
- [90] O. Shahmirzadi, A. Lugowski, and K. Younge, "Text similarity in vector space models: A comparative study," *Proc. - 18th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2019*, pp. 659–666, 2019, doi: 10.1109/ICMLA.2019.00120.
- [91] P. Saipech and P. Seresangtakul, "Automatic Thai Subjective Examination using Cosine Similarity," in *2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA)*, Aug. 2018, pp. 214–218, doi: 10.1109/ICAICTA.2018.8541276.
- [92] G. Somprasertsri, "Content-Based Document Retrieval Using Graph-Based Approach," *J. Technol. Manag. Rajabhat Maha Sarakham Univ.*, vol. 1, no. 7, pp. 77–87, 2020.
- [93] D. Shrestha, S. R. Jeong, and S. Korea, "Digital 7RXULVP ' evelopment and 6ustainability 0 odel for Nepal," pp. 182–187, 2019.
- [94] R. Y. Saputra, L. E. Nugroho, and S. S. Kusumawardani, "Collecting the Tourism Contextual Information data to support the tourism recommendation system," *2019 Int. Conf. Inf. Commun. Technol. ICOIACT 2019*, pp. 79–84, 2019, doi: 10.1109/ICOIACT46704.2019.8938546.
- [95] N. Antonyuk *et al.*, "Consolidated Information Web Resource for Online Tourism Based on Data Integration and Geolocation," *IEEE 2019 14th Int. Sci. Tech. Conf. Comput. Sci. Inf. Technol. CSIT 2019 - Proc.*, vol. 1, pp. 15–20, 2019, doi: 10.1109/STC-CSIT.2019.8929790.
- [96] P. Kotler, "Philip Kotler's contributions to marketing theory and practice.," *Rev. Mark. Res.*, vol. 8, pp. 87–120, 2011, doi: 10.1108/S1548-6435(2011)0000008007.
- [97] Irene Ananda and U. Salamah, "Determination of Sales Data Patterns Using the Association," *Int. J. Inf. Syst. Comput. Sci.*, vol. 2, no. 4, pp. 80–88, 2020.

- [98] R. Muliono, Muhathir, N. Khairina, and M. K. Harahap, "Analysis of Frequent Itemsets Mining Algorithm Againsts Models of Different Datasets," *J. Phys. Conf. Ser.*, vol. 1361, no. 1, 2019, doi: 10.1088/1742-6596/1361/1/012036.
- [99] K. Pitchayadejanant and P. Nakpathom, "Data mining approach for arranging and clustering the agro-tourism activities in orchard," *Kasetsart J. Soc. Sci.*, vol. 39, no. 3, pp. 407–413, 2018, doi: 10.1016/j.kjss.2017.07.004.
- [100] P. Do, & Truong, and H. V Phan, "Developing a BERT based triple classification model using knowledge graph embedding for question answering system," *Appl. Intell.*, vol. 52, pp. 636–651, 2022, doi: 10.1007/s10489-021-02460-w/Published.
- [101] J. Li, Z. Luo, H. Huang, and Z. Ding, "mathematics Towards Knowledge-Based Tourism Chinese Question Answering System," *Z. Towar. Knowledge-Based Tour. Chinese Quest. Answering Syst. Math.*, 2022, doi: 10.3390/math10040664.
- [102] M.-C. Jwa and J.-W. Jwa, "Development of Tourism Information Named Entity Recognition Datasets for the Fine-tune KoBERT-CRF Model," *Int. J. Internet, Broadcast. Commun.*, vol. 14, no. 2, pp. 55–62, 2022, doi: 10.7236/IJIBC.2022.14.2.55.



บรรณานุกรม



ประวัติผู้เขียน

ชื่อ	นางสาวอุมพร ไชยสูง
วันเกิด	วันที่ 15 เดือน ธันวาคม พ.ศ. 2519
สถานที่เกิด	จังหวัดกาฬสินธุ์
สถานที่อยู่ปัจจุบัน	บ้านเลขที่ 202 หมู่ 1 ตำบลหลักเมือง อำเภอภมกลาไสย จังหวัดกาฬสินธุ์ รหัสไปรษณีย์ 46130
ตำแหน่งหน้าที่การงาน	อาจารย์
สถานที่ทำงานปัจจุบัน	สาขาเทคโนโลยีสารสนเทศและมัลติมีเดีย คณะเทคโนโลยีการจัดการ มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสุรินทร์ 145 หมู่ 15 ตำบลนอกเมือง อำเภอเมือง จังหวัดสุรินทร์ รหัสไปรษณีย์ 32000
ประวัติการศึกษา	พ.ศ. 2544 ครุศาสตรบัณฑิต (ค.บ.) โปรแกรมวิชาคอมพิวเตอร์ศึกษา สถาบันราชภัฏจันทรเกษม พ.ศ. 2551 วิทยาศาสตรมหาบัณฑิต (วท.บ.) สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยธุรกิจบัณฑิตย์ พ.ศ. 2565 ปรัชญาดุษฎีบัณฑิต (ปร.ด.) สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยมหาสารคาม
ทุนวิจัย	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน
ผลงานวิจัย	U. Chaisoong and S. Tirakoat, “The Clustering of Questions Affect to Tourist’s Decision Making for Chatbot Design,” 17th Int. Conf. Electr. Eng. Comput. Telecommun. Inf. Technol. ECTI-CON 2020, pp. 784–787, 2020. U. Chaisoong, S. Tirakoat, and C. Jareanpon “Tourist information-seeking behaviours using association rule mining,” ICIC Express Lett., vol. 15, no. 9, pp. 915–923, 2021.