

การทำเหมืองข้อมูลกระดานสนทนาออนไลน์ด้านการท่องเที่ยว โดยใช้กฎความสัมพันธ์ Online Travel Forums Mining using Association Rules

กาญจนา เหล่าเสน (Kanjana Laosen)* และ สุไลมาน สันเกาะ (Sulaimaan Sankoh)*

Received : December 16, 2019

Revised : February 5, 2020

Accepted : March 5, 2020

บทคัดย่อ

ข้อความการท่องเที่ยวบนกระดานสนทนาออนไลน์มีจำนวนมากและเป็นภาษาธรรมชาติทำให้ยากต่อการนำปหาคความรู้ที่เป็นประโยชน์ งานวิจัยนี้ได้นำเสนอแบบจำลองการทำเหมืองข้อมูลกระดานสนทนาออนไลน์ด้านการท่องเที่ยวโดยใช้กฎความสัมพันธ์เพื่อสกัดความรู้ที่อยู่ในข้อความซึ่งประกอบด้วย 5 ขั้นตอน คือ 1) การรวบรวมข้อความกระดานสนทนาออนไลน์ จำนวน 2 ประเภท คือ กระู้และความคิดเห็น 2) การเตรียมข้อความโดยใช้เทคนิคการประมวลผลภาษาธรรมชาติ 3) การวัดความคล้ายคลึงกันของคำเชิงความหมาย 4) การแปลงข้อความสำหรับการหาความสัมพันธ์ 5) การหาความสัมพันธ์ด้วยอัลกอริทึม FP-Growth ผลการทดลองพบว่า 1) ชุดข้อมูลที่เหมาะสมสำหรับการหาความสัมพันธ์ คือ ชุดข้อมูลโดเมนด้านการท่องเที่ยวเนื่องจากสามารถค้นพบกฎความสัมพันธ์ของสถานที่ท่องเที่ยวในช่วงค่าสนับสนุนและค่าความเชื่อมั่นระหว่างร้อยละ 80-100 จำนวนสูงสุดถึง 13 สถานที่ 2) ค้นพบความรู้ที่หลากหลายที่เป็นประโยชน์ต่ออุตสาหกรรมการท่องเที่ยวของจังหวัดภูเก็ต

คำสำคัญ: เหมืองข้อความ กฎความสัมพันธ์การประมวลผลภาษาธรรมชาติ กระดานสนทนาออนไลน์ การวัดความคล้ายคลึงกันของคำเชิงความหมาย

Abstract

There are a lot of online travel forums but they normally are natural language that make difficulty to extract useful knowledge. This study proposes an online travel forums mining

model using association rule for discovering hidden knowledge in travel forums. The proposed model composes of 5 steps; step 1) collecting posts and comments from online travel forums, step 2) pre-processing text using natural language processing technique, step 3) measuring semantic similarity of words by using Porter's algorithm, step 4) transforming text and step 5) discovering association rule by using FP-Growth algorithm. This study found that: 1) applying domain of words for discovering association rule is suitable because it discovers rules that consist of 13 tourist attractions with confidence and support between 80% and 100% 2) there are several useful knowledges obtained from association rules.

Keyword: Text Mining, Association Rule, Natural Language Processing, Online Forum, Word Similarity

1. บทนำ

อุตสาหกรรมการท่องเที่ยวเป็นอุตสาหกรรมหลักที่นำรายได้เข้าสู่ประเทศไทยโดยเฉพาะจังหวัดภูเก็ต ความเข้าใจในพฤติกรรมและความสนใจของนักท่องเที่ยวสามารถทำให้นักท่องเที่ยวได้ มีวิธีการหลายวิธีในการศึกษาพฤติกรรมความสนใจของนักท่องเที่ยวเช่น การใช้แบบสอบถาม การสัมภาษณ์และการสกัดความรู้จากแหล่งข้อมูลขนาดใหญ่กระดานสนทนาออนไลน์ด้านการท่องเที่ยวเป็นช่องทางที่สามารถสะท้อนถึงความสนใจและพฤติกรรมของนักท่องเที่ยวได้เป็นอย่างดี ดังนั้นการค้นหาคำความรู้จากกระดานสนทนาออนไลน์จึงเป็นช่องทางที่สามารถค้นพบความสนใจและพฤติกรรมของนักท่องเที่ยวในแต่ละกลุ่มได้ [1]

* วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตภูเก็ต

* College of Computing, Prince of Songkla University, Phuket Campus

งานวิจัยนี้จึงนำเสนอแบบจำลองการทำเหมืองข้อมูลจากกระดานสนทนาออนไลน์ด้านการท่องเที่ยวโดยใช้กฎความสัมพันธ์สำหรับค้นหาความสนใจและพฤติกรรมของนักท่องเที่ยวเพื่อเป็นองค์ความรู้ให้แก่ผู้ประกอบการด้านการท่องเที่ยวสำหรับพัฒนาสินค้าและบริการให้ตรงตามกลุ่มเป้าหมายนักท่องเที่ยวในแต่ละกลุ่มประเทศ โดยงานวิจัยนี้เลือกกลุ่มตัวอย่างข้อความด้านการท่องเที่ยวที่เกี่ยวข้องกับจังหวัดภูเก็ต เป็นข้อความภาษาธรรมชาติและใช้ภาษาอังกฤษซึ่งเป็นภาษาหลักที่ใช้ในการสื่อสารของคนทั่วโลกทำให้ได้รับมุมมองความคิดเห็นที่หลากหลายจากคนทั่วโลกเช่นกัน

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้ได้ศึกษาทฤษฎีและงานวิจัยสำหรับการออกแบบแบบจำลองการทำเหมืองข้อมูลกระดานสนทนาออนไลน์ด้านการท่องเที่ยวโดยมีรายละเอียดดังนี้

2.1 กระดานสนทนาออนไลน์

กระดานสนทนาออนไลน์ (Online Forums) คือเว็บแอปพลิเคชันสำหรับจัดเก็บเนื้อหา กระทู้ และการอภิปรายที่ผู้ใช้งานอินเทอร์เน็ตสร้างขึ้นโดยมีโดเมนเฉพาะ [2] เช่น เศรษฐกิจ การเมือง และการท่องเที่ยว กระดานสนทนาออนไลน์ด้านการท่องเที่ยวเป็นสื่อสังคมออนไลน์ประเภทหนึ่งสำหรับตั้งกระทู้และอภิปรายด้านการท่องเที่ยวจากผู้ที่มีความสนใจด้านการท่องเที่ยว เช่น “Rough Guides” “Travellerspoint” และ “TripAdvisor” กระดานสนทนาออนไลน์ TripAdvisor เป็นส่วนหนึ่งของเว็บไซต์ TripAdvisor ซึ่งเป็นเว็บไซต์ที่มีจำนวนผู้เขียนคำวิจารณ์และแสดงความคิดเห็นด้านการท่องเที่ยวเป็นจำนวนมาก [3] นอกจากนี้ข้อความในกระดานสนทนาออนไลน์จำนวนมากเหล่านั้นสามารถนำมาสกัดความรู้ที่เกี่ยวข้องกับพฤติกรรมของนักท่องเที่ยวด้วยงานวิจัยนี้จึงได้รวบรวมข้อความการสนทนาจากกระดานสนทนาออนไลน์ของ TripAdvisor

2.2 การท่องเที่ยว

การท่องเที่ยว (Tourism) เป็นกิจกรรมของบุคคลที่เดินทางและอยู่ในสถานที่นอกสภาพแวดล้อมปกติเพื่อการพักผ่อน ธุรกิจ และวัตถุประสงค์อื่น ๆ [4] ข้อมูลและงานวิจัยที่เกี่ยวข้องกับการท่องเที่ยวมีดังนี้

2.2.1 ประเภทนักท่องเที่ยว กระทรวงการท่องเที่ยวและกีฬาได้จัดเก็บข้อมูลของนักท่องเที่ยวต่างชาติโดยใช้

สัญชาตินักท่องเที่ยวและแบ่งกลุ่มนักท่องเที่ยวตามภูมิศาสตร์ [5] และงานวิจัยนี้จึงได้จัดกลุ่มนักท่องเที่ยวเป็น 4 กลุ่มตามภูมิศาสตร์ ประกอบด้วย เอเชีย ยุโรป โอเชียเนีย และอเมริกา การจัดกลุ่มของนักท่องเที่ยวนี้เพื่อต้องการทราบถึงพฤติกรรมและปัจจัยที่มีผลต่อการเดินทางมาท่องเที่ยวของนักท่องเที่ยวในแต่ละกลุ่ม

2.2.2 สถานที่ท่องเที่ยวในจังหวัดภูเก็ต จากหนังสือแนะนำการท่องเที่ยวจังหวัดภูเก็ตระบุสถานที่ท่องเที่ยวในจังหวัดภูเก็ตทั้งหมดมีจำนวน 63 สถานที่ [6] แต่มีเพียงจำนวน 55 สถานที่เท่านั้นที่ปรากฏบนกระดานสนทนาออนไลน์ TripAdvisor และงานวิจัยนี้ใช้แหล่งข้อความจากกระดานสนทนาออนไลน์ TripAdvisor จึงเลือกใช้เฉพาะ 55 สถานที่เท่านั้น

2.2.3 โดเมนด้านการท่องเที่ยว งานวิจัยนี้ได้ประยุกต์โดเมนด้านการท่องเที่ยวจากงานวิจัยที่กำหนดโดเมนด้านการท่องเที่ยวของประเทศไทย [7] -[10] โดยเลือกโดเมนที่แสดงถึงพฤติกรรมและปัจจัยที่มีผลต่อการเดินทางมาท่องเที่ยวในจังหวัดภูเก็ตของนักท่องเที่ยว ประกอบด้วย 8 โดเมน ได้แก่ “Accommodation” “Activity” “Contact” “Food” “Infrastructure” “Location” “Time” และ “Payment” เพื่อใช้เป็นชุดข้อมูล

2.3 การประมวลผลภาษาธรรมชาติ

การประมวลผลภาษาธรรมชาติเป็นศาสตร์ที่ศึกษาค้นคว้า วิจัยวิธีที่ทำให้คอมพิวเตอร์สามารถเข้าใจภาษาธรรมชาติหรือภาษาพูดของมนุษย์เพื่อนำไปใช้ประโยชน์ [11] งานวิจัยนี้ได้ใช้ทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการประมวลผลภาษาธรรมชาติมีดังนี้

2.3.1 เวิร์ดเน็ต (WordNet) [12] เป็นฐานข้อมูลคำศัพท์ภาษาอังกฤษที่จัดหมวดหมู่คำที่มีความหมายเหมือนกันไว้ด้วยกันเป็นกลุ่มซึ่งมีคำอธิบายความหมายและการเชื่อมโยงความสัมพันธ์เชิงความหมายระหว่างคำฐานข้อมูลเวิร์ดเน็ตถูกนำไปใช้อย่างแพร่หลายในงานวิจัยที่เกี่ยวข้องกับการจัดกลุ่มคำหรือข้อความและการวัดความคล้ายคลึงกันเชิงความหมาย เช่น การประเมินความคล้ายคลึงกันระหว่างข้อความสั้น ๆ และประโยค [13] งานวิจัยนี้ต้องการการจัดกลุ่มคำที่มีความหมายคล้ายคลึงกันจึงเลือกใช้ฐานข้อมูลเวิร์ดเน็ต

2.3.2 การวัดความคล้ายคลึงกันของคำเชิงความหมาย

(Semantic Similarity of Words) เป็นการวัดความคล้ายคลึงกันระหว่างคำสองคำ งานวิจัยนี้ได้เลือกใช้การวัดโดยใช้ความยาวของเส้นทางที่เกิดร่วมกันของคำ 2 คำและความลึกของแต่ละคำที่สั้นที่สุด (Wu and Palmer's Measure) [14] วิธีนี้คำนวณตำแหน่งของคำและเส้นทางที่ปรากฏร่วมกันระหว่างคำในลำดับชั้นฐานข้อมูลเวกเตอร์เน็ตซึ่งความยาวของลำดับชั้นที่ปรากฏร่วมกันจะแสดงถึงความคล้ายคลึงกันเชิงความหมายระหว่างคำ โดยคำนวณได้จากสมการที่ (1)

$$sim(w1, w2) = \frac{2 * depth(lso(w1, w2))}{depth(w1) + depth(w2)} \quad (1)$$

โดยที่

$w1, w2$	คือ คำที่ 1 และ คำที่ 2
sim	คือ ความคล้ายคลึงกันระหว่างคำ
$depth(lso)$	คือ ความลึกที่เกิดร่วมกัน
$depth$	คือ ความลึกของคำ

2.3.3 คำหยุด (Stopword) เป็นคำที่เกิดขึ้นบ่อยในเอกสารแต่ไม่มีความหมายต่อเนื้อหา [15] การกำจัดคำหยุดทำให้ประหยัดพื้นที่ของหน่วยความจำและเวลาสำหรับการประมวลผล [16] ตัวอย่างคำหยุดในข้อความภาษาอังกฤษ เช่น “a” “an” “at” “for” และ “of” งานวิจัยนี้ใช้ข้อความภาษาอังกฤษที่มีคำหยุดจำนวนมาก จึงกำจัดคำหยุดเพื่อลดคำที่ไม่มีความสำคัญสำหรับการนำคำไปใช้หากฎความสัมพันธ์

2.3.4 การหารากศัพท์ของคำ (Stemming) เป็นกระบวนการแปลงรูปแบบต่าง ๆ ของคำให้กระชับและแม่นยำ [17] เนื่องจากคำในภาษาอังกฤษบางคำมีความหมายในรากศัพท์เดียวกันแต่มีรูปแบบการเขียนที่ต่างกัน เช่น คำว่า “travel” “traveling” และ “travelling” ดังนั้นการแปลงไปยังรากศัพท์ของคำจึงสามารถลดรูปแบบทางไวยากรณ์หรือคำที่แตกต่างกันได้ งานวิจัยนี้ได้รวบรวมข้อความจากสื่อสังคมออนไลน์จึงเลือกใช้ Porters Stemmer [18] สำหรับการค้นหารากศัพท์ของคำภาษาอังกฤษเพื่อเตรียมคำสำหรับการวัดความคล้ายคลึงกันของคำเชิงความหมาย

2.4 การทำเหมืองข้อมูล

การทำเหมืองข้อมูลเป็นกระบวนการวิเคราะห์เพื่อหารูปแบบกฎ หรือความรู้จากข้อมูลจำนวนมาก [19] งานวิจัยนี้ได้ใช้ทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการทำเหมืองข้อมูลมีดังนี้

2.4.1 การหากฎความสัมพันธ์

กฎความสัมพันธ์ (Association Rule) เป็นรูปแบบความสัมพันธ์ของชุดข้อมูลที่ปรากฏขึ้นในฐานข้อมูล [19] ลักษณะของกฎความสัมพันธ์อยู่ในรูปของ IF-THEN โดยที่ IF (ถ้า) แทนเงื่อนไขและ THEN (แล้ว) แทนข้อสรุป ขั้นตอนหากฎความสัมพันธ์ [19] มี 3 ขั้นตอนดังนี้

$$Support(A \Rightarrow B) = P(A \cup B) \quad (2)$$

1) การคำนวณค่าสนับสนุน (Support Value)

เป็นร้อยละของชุดข้อมูลที่ปรากฏในฐานข้อมูลโดยคำนวณได้จากสมการ (2)

โดยที่

A, B	คือ ชุดข้อมูล
$Support$	คือ ค่าสนับสนุน
$P(A \cup B)$	คือ ชุดข้อมูล A และ B ปรากฏพร้อมกัน

$$Confidence(A \Rightarrow B) = \frac{Support_count(A \cup B)}{Support_count(A)} \quad (3)$$

โดยที่

A, B	คือ ชุดข้อมูล
$Confidence$	คือ ค่าความเชื่อมั่น
$Support_count(AB)$	คือ ค่าสนับสนุนที่ปรากฏ A และ B พร้อมกัน
$Support_count(A)$	คือ ค่าสนับสนุนที่ปรากฏ A

2) การคำนวณค่าความเชื่อมั่น (Confidence Value)

เป็นร้อยละของกฎที่พบเมื่อปรากฏ A แล้วจะปรากฏ B พร้อมกันโดยคำนวณได้จากสมการ (3)

$$Lift(A \Rightarrow B) = \frac{P(A \cup B)}{P(A)P(B)} \quad (4)$$

โดยที่

A, B	คือ ชุดข้อมูล
$Lift$	คือ ค่าลิฟท์
$P(AB)$	คือ ปรากฏชุดข้อมูล A และ B พร้อมกัน
$P(A)$	คือ ปรากฏชุดข้อมูล A
$P(B)$	คือ ปรากฏชุดข้อมูล B

3) การคำนวณค่าลิฟท์ (Lift Value) เป็นร้อยละของจำนวนชุดข้อมูลของกฎที่ปรากฏในฐานข้อมูลต่อจำนวนชุดข้อมูลที่ปรากฏ A และ B โดยคำนวณได้จากสมการ (4)

2.4.2 อัลกอริทึมการหากฎความสัมพันธ์

อัลกอริทึมการหากฎความสัมพันธ์ที่ถูกนำมาใช้ในงานวิจัยอย่างแพร่หลาย ได้แก่ Apriori และ FP-Growth โดยอัลกอริทึม Apriori [20] เป็นอัลกอริทึมที่ทำงานได้ดีในฐานข้อมูลขนาดเล็กแต่ต้องอ่านข้อมูลจากฐานข้อมูลหลายครั้งเพื่อสร้างกลุ่มข้อมูลผู้ทำซึ่งทำให้จำเป็นต้องใช้หน่วยความจำสำหรับประมวลผลจำนวนมากและในกรณีที่มิชุดข้อมูลจำนวนมากจะต้องใช้เวลาสำหรับอ่านชุดข้อมูลนาน ส่วนอัลกอริทึม FP-Growth อ่านข้อมูลจากฐานข้อมูลเพียง 2 ครั้งและไม่สร้างกลุ่มข้อมูลผู้ทำซึ่งทำให้ใช้เวลาและหน่วยความจำน้อยกว่าอัลกอริทึม Apriori [21] งานวิจัยนี้มีฐานข้อมูลที่มีขนาดใหญ่จึงเลือกอัลกอริทึมที่สามารถรับรองฐานข้อมูลของงานวิจัยนี้ได้คืออัลกอริทึม FP-Growth

3. แบบจำลองการทำเหมืองข้อมูลกระดานสนทนาออนไลน์ด้านการท่องเที่ยวโดยใช้กฎความสัมพันธ์

งานวิจัยนี้นำเสนอแบบจำลองการทำเหมืองข้อมูลการท่องเที่ยวจากกระดานสนทนาออนไลน์ด้านการท่องเที่ยวโดยใช้กฎความสัมพันธ์ ซึ่งได้แบ่งการทดลองออกเป็น 2 การทดลอง คือ 1) การทดลอง A หากฎความสัมพันธ์ระหว่างสถานที่ท่องเที่ยวและคำ และ 2) การทดลอง B หากฎความสัมพันธ์ระหว่างสถานที่ท่องเที่ยวและโดเมนด้านการท่องเที่ยวโดยแสดงดังภาพที่ 1 ซึ่งประกอบด้วย 5 ขั้นตอนดังนี้

3.1 ขั้นตอนที่ 1

การรวบรวมข้อความจากกระดานสนทนาออนไลน์ TripAdvisor ที่เกี่ยวกับการท่องเที่ยวจังหวัดภูเก็ต ประกอบด้วย 2 ประเภท คือ กระู้ (Post) และความคิดเห็น (Comment) แสดงตัวอย่างข้อความดังภาพที่ 2 โดยรูป A) กระู้เป็นข้อความอธิบายคำถามหรือปัญหาด้านการท่องเที่ยว และรูป B) ความคิดเห็นเป็นคำตอบหรือข้อคิดเห็นจากผู้ร่วมสนทนาบนกระดานสนทนาออนไลน์ผลลัพธ์จากขั้นตอนนี้ คือ กระู้ที่มีข้อความของนักท่องเที่ยวเอเชียจำนวน 2,736 ข้อความ โอเชียเนียจำนวน 8,421 ข้อความ ยุโรปจำนวน 2,576 ข้อความ และอเมริกาจำนวน 381

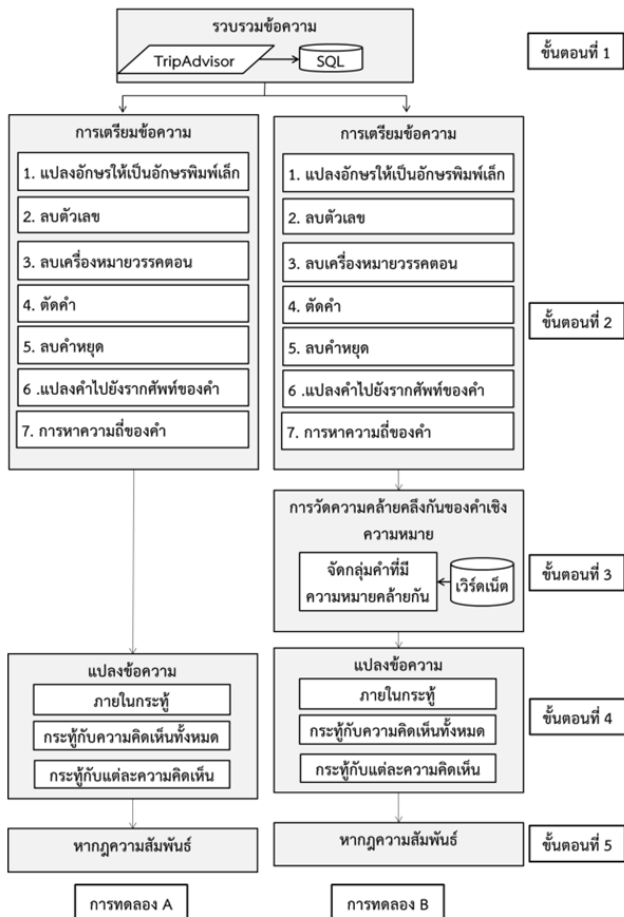
ข้อความ ส่วนความคิดเห็นมีข้อความของนักท่องเที่ยวเอเชียจำนวน 3,355 ข้อความ โอเชียเนียจำนวน 37,358 ข้อความ ยุโรปจำนวน 3,978 ข้อความ และอเมริกาจำนวน 378 ข้อความ

3.2 ขั้นตอนที่ 2

การเตรียมข้อความประกอบด้วยขั้นตอนย่อยจำนวน 7 ขั้นตอน ได้แก่ 1) แปลงอักขรพิมพ์ใหญ่ให้เป็นอักขรพิมพ์เล็ก 2) ลบตัวเลข 3) ลบเครื่องหมายวรรคตอน 4) ตัดคำเพื่อตัดข้อความให้อยู่ในรูปแบบของคำ 5) ลบคำหยุดโดยใช้รายการคำหยุดของ Package ‘tm’ ในโปรแกรม R 6) แปลงคำไปยังรากศัพท์ และ 7) นับความถี่ของคำเพื่อเลือกคำที่มีค่าความถี่ที่สามารถนำไปใช้หากฎความสัมพันธ์และวัดความคล้ายคลึงกันของคำเชิงความหมายในขั้นตอนถัดไป ตัวอย่างการเตรียมข้อความนำเข้า “Tour suggestions staying Karon for 7 nights- any suggestions” แสดงดังตารางที่ 1 ผลลัพธ์จากขั้นตอนนี้ คือคำจากกระู้และความคิดเห็นจำนวน 335 คำ และ 256 คำตามลำดับ

3.3 ขั้นตอนที่ 3

การวัดความคล้ายคลึงกันของคำเชิงความหมายขั้นตอนนี้ใช้เฉพาะการทดลอง B เท่านั้นเพื่อจัดกลุ่มคำที่ได้จากการเตรียมข้อความให้อยู่ในโดเมนด้านการท่องเที่ยว การหาค่าความคล้ายคลึงกันของคำเชิงความหมายระหว่างคำและโดเมนด้านการท่องเที่ยวโดยนำคำจากกระู้และความคิดเห็นจับคู่กับโดเมนด้านการท่องเที่ยวทั้ง 8 โดเมนด้วยวิธีการของ Wu and Palmer ตัวอย่างผลลัพธ์ของการจัดคำให้อยู่ในโดเมนแสดงดังตารางที่ 2 จากตัวอย่างตารางที่ 2 ค่าความคล้ายคลึงกันของคำเชิงความหมายระหว่างคำว่า “hotel” และ “Accommodation” มีค่าเท่ากับ 0.74 และระหว่างคำว่า “hotel” และ “Activity” มีค่าเท่ากับ 0.43 หลังจากนั้นนำค่าความคล้ายคลึงกันมาเปรียบเทียบเพื่อจัดกลุ่มคำตามโดเมนด้านการท่องเที่ยวโดยเลือกจากคำที่มีค่ามากที่สุด ตัวอย่าง คำว่า “hotel” เมื่อเปรียบเทียบกับทุกโดเมนแล้ว โดเมน “Accommodation” มีค่ามากที่สุดซึ่งมีค่า 0.74 ดังนั้นคำว่า “hotel” จึงถูกจัดอยู่ในกลุ่มโดเมน “Accommodation” โดยจะเปรียบเทียบจนครบทุกคำจึงได้คำในแต่ละโดเมนทั้งหมด

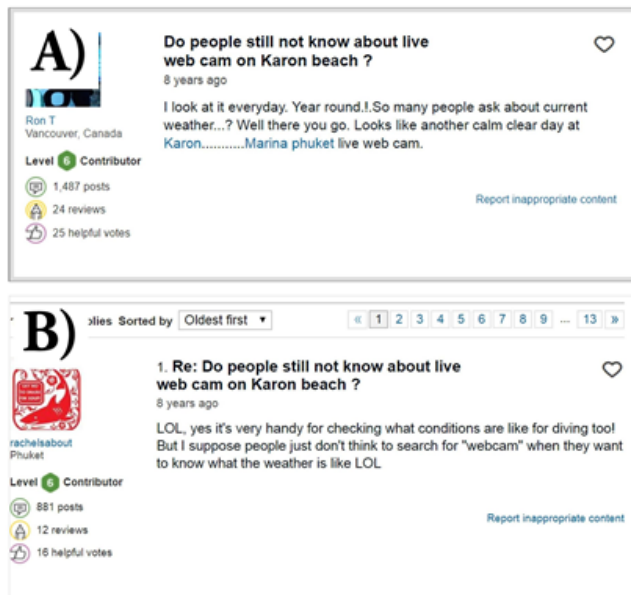


ภาพที่ 1 แบบจำลองการทำเหมืองข้อมูล กระดานสนทนาออนไลน์ด้านการท่องเที่ยวโดยใช้กฎความสัมพันธ์

3.4 ขั้นตอนที่ 4

การแปลงข้อความให้อยู่ในรูปแบบตารางค่า T (True) และ F (False) โดยที่ T หมายถึงปรากฏข้อความที่กำหนดและ F หมายถึงไม่ปรากฏข้อความที่กำหนดสำหรับหากฎความสัมพันธ์สามารถแบ่งการแปลงข้อความได้ดังนี้ การทดลอง A แปลงข้อความโดยใช้คำ และการทดลอง B แปลงข้อความโดยใช้โดเมนการท่องเที่ยว โดยการทดลอง A และ B แบ่งการทดลองย่อยออกเป็น 3 รูปแบบ คือ แบบที่ 1: การแปลงข้อความภายในกระทู้ (A-1) และ (B-1) เพื่อแสดงความสัมพันธ์ภายในคำถาม แบบที่ 2: การแปลงข้อความระหว่างกระทู้และความคิดเห็นทั้งหมดของกระทู้ (A-2) และ (B-2) เพื่อแสดงความสัมพันธ์ระหว่างคำถามและคำตอบของผู้เข้ามาแสดงความคิดเห็นทั้งหมด

แบบที่ 3 : การแปลงข้อความระหว่างกระทู้และแต่ละความคิดเห็น (A-3) และ (B-3) แสดงความสัมพันธ์ระหว่างคำถามและคำตอบของแต่ละคน ตัวอย่างการแปลงข้อความดังภาพที่ 3 เป็นการแปลงข้อความภายในกระทู้ โดยกำหนดให้วงกลมสีขาวแสดงชื่อสถานที่ท่องเที่ยวและวงกลมสีเทาแสดงคำที่ได้จากการเตรียมข้อความกระทู้ ตารางพื้นหลังสีขาวแสดงชุดข้อมูลชื่อสถานที่ท่องเที่ยวและตารางพื้นหลังสีเทาแสดงชุดข้อมูลคำ ดังนั้นจำนวนระเบียบที่ได้จากการเตรียมข้อความของแต่ละการทดลองแสดงดังตารางที่ 3



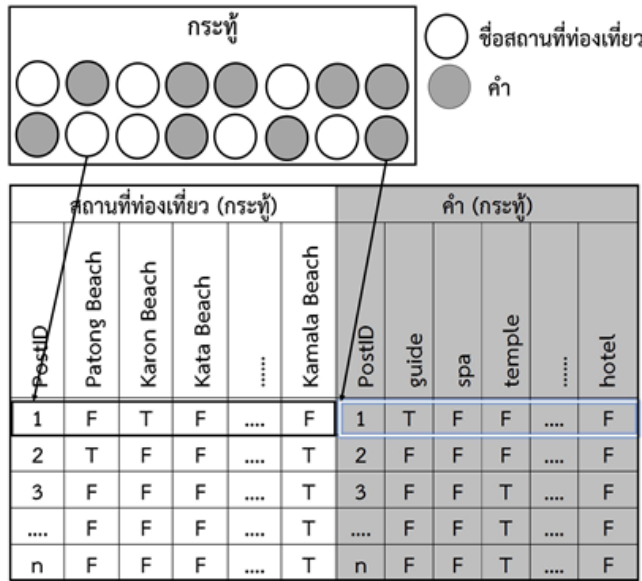
ภาพที่ 2 A) กระทู้ และ B) ความคิดเห็น

3.5 ขั้นตอนที่ 5

การหากฎความสัมพันธ์โดยนำตารางค่า T และ F ไปหากฎความสัมพันธ์ด้วยอัลกอริทึม FP-Growth โดยกำหนดค่าสนับสนุนขั้นต่ำร้อยละ 20 ค่าความเชื่อมั่นขั้นต่ำร้อยละ 50 และค่าลิฟท์มากกว่า 1

4. ผลการทดลอง

งานวิจัยนี้จัดกลุ่มกฎความสัมพันธ์เพื่อใช้เปรียบเทียบสำหรับการค้นหาความรู้โดยแบ่งกลุ่มตามค่าความเชื่อมั่นออกเป็น 3 กลุ่ม คือ ช่วงค่าความเชื่อมั่นร้อยละ 50-59 ร้อยละ 60-79 และร้อยละ 80-100 และในแต่ละช่วงค่าความเชื่อมั่นได้ แบ่งเป็น 3 กลุ่มตามค่าสนับสนุน คือ



ภาพที่ 3 ตัวอย่างการแปลงข้อความของการทดลอง A-1

ตารางที่ 1 คำและความถี่หลังจากการเตรียมข้อความ

คำ	ความถี่ (ครั้ง)
tour	1
suggest	2
stay	1
karon	1
night	1

ช่วงค่าสนับสนุนร้อยละ 20-49 ร้อยละ 50-79 และร้อยละ 80-100 ในบทความนี้ผู้วิจัยได้แสดงกฎความสัมพันธ์เฉพาะช่วงค่าความเชื่อมั่นและค่าสนับสนุนร้อยละ 80-100 เท่านั้น เนื่องจากเป็นกฎความสัมพันธ์ที่สามารถแสดงความถี่และบ่งบอกถึงความสัมพันธ์ของชุดข้อมูลที่สูงที่สุดของการทดลองนี้ซึ่งจะเป็นประโยชน์ต่อการนำไปใช้ในอุตสาหกรรมการท่องเที่ยวของจังหวัดภูเก็ตต่อไปโดยมีรายละเอียดดังนี้

4.1 ผลการทดลอง A

กฎความสัมพันธ์ระหว่างสถานที่ท่องเที่ยวและคำที่ค่าสนับสนุนและค่าความเชื่อมั่นระหว่างร้อยละ 80-100 แสดงตัวอย่างดังตารางที่ 4 กฎข้อที่ 1 หมายความว่า สำหรับ นักท่องเที่ยว กลุ่มประเทศ โอเชียเนีย ถ้ากระตุ้มีการกล่าวถึง “Kamala Beach” แล้วแต่จะความคิดเห็นจะกล่าวถึงคำว่า “idea” เสมอด้วยค่าสนับสนุนและค่าความเชื่อมั่นระหว่างร้อยละ 80-100 ผลการทดลอง A

ตารางที่ 2 ค่าความคล้ายคลึงกันของค่าเชิงความหมายระหว่างคำกับโดเมน

คำ	โดเมนการท่องเที่ยว							
	accommodation	activity	time	location	Payment	infrastructure	contact	food
hotel	0.74	0.43	0.29	0.57	0.24	0.24	0.63	0.40
massage	0.70	0.78	0.53	0.74	0.60	0.32	0.63	0.44
afternoon	0.59	0.43	0.80	0.38	0.38	0.38	0.40	0.40
village	0.42	0.43	0.46	0.71	0.38	0.38	0.42	0.40
rent	0.73	0.63	0.57	0.77	0.82	0.77	0.67	0.43
base	0.70	0.55	0.57	0.91	0.53	1.00	0.76	0.67
email	0.63	0.71	0.62	0.67	0.63	0.33	0.78	0.47
seafood	0.35	0.50	0.33	0.50	0.27	0.27	0.38	0.92

แสดงให้เห็นรายละเอียดดังนี้

1) การทดลอง A-1 และ A-2 ไม่พบกฎความสัมพันธ์ เนื่องจากการทดลองดังกล่าวมีค่าสนับสนุนและค่าความเชื่อมั่นต่ำกว่าร้อยละ 80-100

2) สถานที่ท่องเที่ยวที่มีค่าสนับสนุนและค่าความเชื่อมั่นมากกว่าสถานที่ท่องเที่ยวอื่น ๆ คือ หาดกมลาซึ่งเป็นกฎความสัมพันธ์จากกลุ่มนักท่องเที่ยวโอเชียเนียและอเมริกา

3) กฎความสัมพันธ์กลุ่มนักท่องเที่ยวอเมริกาและโอเชียเนียปรากฏสถานที่ท่องเที่ยวและคำที่เหมือนกัน นอกจากนี้มีช่วงค่าสนับสนุนและค่าความเชื่อมั่นของนักท่องเที่ยวทั้ง 2 กลุ่มอยู่ในช่วงเดียวกันยกเว้นกฎความสัมพันธ์ข้อที่ 5

4) การทดลอง A-2 และ A-3 เป็นการทดลองที่ใช้แหล่งข้อมูลจากกระตุ้และความคิดเห็นเหมือนกันแต่มีวิธีการแปลงข้อความที่ต่างกันส่งผลให้ได้ผลการทดลองที่ 80-100 เท่านั้นที่สามารถค้นหากฎความสัมพันธ์ที่มีค่าสนับสนุนและค่าความเชื่อมั่นในช่วงร้อยละ 80-100 ได้ขณะที่การทดลอง A-2 ไม่พบกฎความสัมพันธ์ในช่วงค่าสนับสนุนและค่าความเชื่อมั่นดังกล่าว ส่วนการทดลอง A-1 เป็นชุดข้อมูลจากกระตุ้เท่านั้น

4.2 ผลการทดลอง B

กฎความสัมพันธ์ระหว่างสถานที่ท่องเที่ยวและโดเมนด้านการท่องเที่ยวที่ค่าสนับสนุนและค่าความเชื่อมั่นระหว่างร้อยละ 80-100 แสดงตัวอย่างดังตารางที่ 5 กฎข้อที่ 1 หมายความว่า สำหรับนักท่องเที่ยวกลุ่มประเทศเอเชียเนี่ย ถ้ากระตุ้กล่าวถึง “Panwa Cape” แล้วแต่ละความคิดเห็นจะกล่าวถึงคำที่อยู่ในโดเมน “Accommodation” เสมอด้วยค่าสนับสนุนและค่าความเชื่อมั่นระหว่างร้อยละ 80-100 ผลการทดลอง B แสดงให้เห็นรายละเอียดดังนี้

1) เมื่อพิจารณาค่าสนับสนุนและค่าความเชื่อมั่นระหว่างร้อยละ 80-100 พบว่าไม่พบกฎความสัมพันธ์ของการทดลอง B-2 เนื่องจากการทดลองดังกล่าวมีค่าสนับสนุนและค่าความเชื่อมั่นต่ำกว่าร้อยละ 80-100

2) แหล่มพันว่า หาดกมลา พระพุทธมิ่งมงคลเอกนาคคีรี และหาดในหานเป็นสถานที่ท่องเที่ยวที่มีค่าสนับสนุนและมีค่าความเชื่อมั่นในช่วงร้อยละ 80-100 ซึ่งเป็นช่วงค่าสนับสนุนและช่วงค่าความเชื่อมั่นที่สูงที่สุด

3) โดเมน “Accommodation” “Time” และ “Contact” เป็นโดเมนด้านการท่องเที่ยวที่มีค่าสนับสนุนและค่าความเชื่อมั่นมากในช่วงร้อยละ 80-100 ซึ่งเป็นโดเมนด้านการท่องเที่ยวที่มีค่าสนับสนุนและค่าความเชื่อมั่นสูงสุดของการทดลอง B

4) การทดลอง B-2 และ B-3 เป็นการทดลองที่ใช้แหล่งข้อมูลจากความคิดเห็นเหมือนกันแต่มีวิธีการแปลงข้อความที่ต่างกันทำให้ได้ผลการทดลองที่ต่างกันด้วยผลลัพธ์ของการทดลอง B-3 กฎความสัมพันธ์โดยส่วนใหญ่มีช่วงค่าสนับสนุนที่สูงกว่าการทดลอง B-2

4.3 การเปรียบเทียบกฎความสัมพันธ์

การเปรียบเทียบจำนวนกฎความสัมพันธ์จากการทดลอง A และ B มีรายละเอียดดังนี้

1) การทดลอง B ค้นพบสถานที่ท่องเที่ยวมากกว่าการทดลอง A ดังตารางที่ 6 เนื่องจากการทดลอง A หากกฎความสัมพันธ์โดยใช้ชุดข้อมูลคำที่มีจำนวนมากและไม่ได้จัดกลุ่มคำทำให้ความถี่ของการค้นหาชุดข้อมูลที่ปรากฏขึ้นบ่อยในชุดข้อมูลกระจายไปตามจำนวนคำทำให้ความถี่ค่าสนับสนุนน้อยและไม่ถึงค่าสนับสนุนขั้นต่ำที่กำหนดจึงได้กฎที่มีจำนวนสถานที่ท่องเที่ยวที่น้อย ส่วนการทดลอง B หากกฎความสัมพันธ์โดยใช้ชุดข้อมูลโดเมนด้านการท่องเที่ยว

ซึ่งโดเมนด้านการท่องเที่ยวมาจากคำจำนวนหลายคำที่มีความหมายคล้ายคลึงกันทำให้ความถี่ของชุดข้อมูลโดเมนด้านการท่องเที่ยวมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำที่กำหนดจึงได้กฎที่มีจำนวนสถานที่ท่องเที่ยวจำนวนมาก

2) ความคล้ายคลึงกันของกฎความสัมพันธ์จากตารางที่ 7 พบว่าการทดลอง A ได้กฎความสัมพันธ์ของคำที่มีความคล้ายคลึงกันของคำเชิงความหมาย เช่น กฎข้อที่ 1 สถานที่ท่องเที่ยว “Patong Beach” และคำว่า “price” และกฎข้อที่ 2 สถานที่ท่องเที่ยว “Patong Beach” และคำว่า “rent” ซึ่งคำว่า “price” และ “rent” เมื่อพิจารณาความคล้ายคลึงกันของคำเชิงความหมายแล้วคำทั้ง 2 คำสามารถอยู่ในโดเมนเดียวกันได้คือ โดเมน “Payment” ซึ่งสอดคล้องกับการทดลอง B ซึ่งได้กฎความสัมพันธ์ของโดเมนด้านการท่องเที่ยวเช่น กฎข้อที่ 3 สถานที่ท่องเที่ยว “Patong Beach” และโดเมน “Payment”

4.4 การเปรียบเทียบงานวิจัยนี้กับงานวิจัยที่เกี่ยวข้องและวิจารณ์

งานวิจัยที่เกี่ยวข้องที่ใช้เทคนิคการหาความสัมพันธ์ค้นหาความรู้จากข้อความแสดงดังตารางที่ 8 ซึ่งแสดงให้เห็นถึงความแตกต่างของงานวิจัยดังนี้

1) งานวิจัยลำดับที่ 1-2 เป็นงานวิจัยที่ไม่ได้นำวิธีการประมวลผลภาษาธรรมชาติมาใช้ทำให้งานวิจัยลำดับที่ 1 กำหนดค่าสนับสนุนขั้นต่ำได้ต่ำมาก ส่วนงานวิจัยลำดับที่ 3-5 เป็นงานวิจัยที่นำวิธีการประมวลผลภาษาธรรมชาติมาใช้ทำให้สามารถกำหนดค่าสนับสนุนขั้นต่ำได้สูง

เนื่องจากการประมวลผลภาษาธรรมชาติสามารถช่วยให้อธิบายข้อความในชุดข้อมูลมีความถูกต้องมากยิ่งขึ้น เช่น การลบคำหยุดจะค้นหาความถี่ของชุดข้อมูลโดยคำนวณเฉพาะคำที่มีความสำคัญและมีผลต่องานวิจัยเท่านั้นจึงส่งผลให้ได้ค่าสนับสนุนของชุดข้อมูลที่สูงขึ้น อีกทั้งการแปลงคำไปยังรากศัพท์ของคำจะสามารถนับรวมคำที่มีความหมายเหมือนกันแต่เขียนต่างกันได้ซึ่งจะส่งผลให้ได้ค่าสนับสนุนของชุดข้อมูลที่สูงขึ้นและมีความถูกต้องมากยิ่งขึ้น งานวิจัยนี้ได้ดำเนินการประมวลผลภาษาธรรมชาติใช้ในการเตรียมข้อความ

2) งานวิจัยนี้ใช้ชุดข้อมูลที่แตกต่างจากงานวิจัยที่เกี่ยวข้อง โดยงานวิจัยที่เกี่ยวข้องจะใช้ชุดข้อมูลโดยตรงจากข้อความ



ตารางที่ 3 จำนวนระเบียบที่ได้จากการเตรียมข้อความ

กลุ่มประเทศ	จำนวนระเบียบทั้งหมด	จำนวนระเบียบของแต่ละการทดลอง					
		A-1	A-2	A-3	B-1	B-2	B-3
เอเชีย	2,736	1,119	596	1,303	421	128	300
โอเชียเนีย	8,421	3,645	2,865	14,976	1,436	611	3,203
ยุโรป	2,570	1,098	645	1,564	373	124	299
อเมริกา	381	125	59	127	45	11	23

ตารางที่ 4 กฎความสัมพันธ์ของการทดลอง A

ที่	สถานที่ท่องเที่ยว	โดเมนการท่องเที่ยว	สนับสนุน	ความเชื่อมั่น	ลิฟท์	การทดลอง	กลุ่มประเทศ
1	Panwa Cape	Accommodation	0.88	0.91	1.01	B-3	โอเชียเนีย
2	Kamala Beach	Contact	0.87	1	1.05	B-3	อเมริกา
3	Panwa Cape	Time	0.84	0.88	1.01	B-1	เอเชีย
4	Big Buddha	Time	0.83	0.88	1.01	B-1	เอเชีย
5	Nai Harn Beach	Contact	0.81	0.82	1.01	B-1	เอเชีย
6	Panwa Cape	Location, Accommodation	0.8	0.82	1.01	B-3	โอเชียเนีย

ตารางที่ 5 กฎความสัมพันธ์ของการทดลอง B

ที่	สถานที่ท่องเที่ยว	คำ	สนับสนุน	ความเชื่อมั่น	ลิฟท์	การทดลอง	กลุ่มประเทศ
1	Kamala Beach	idea	0.9	0.95	1.01	A-3	โอเชียเนีย
2	Kamala Beach	idea	0.9	0.95	1.01	A-3	อเมริกา
3	Kamala Beach	trip	0.87	0.93	1.01	A-3	โอเชียเนีย
4	Kamala Beach	trip	0.87	0.93	1.01	A-3	อเมริกา
5	Kamala Beach	idea, check	0.85	0.9	1.01	A-3	โอเชียเนีย
6	Kamala Beach	idea, trip	0.83	0.88	1.02	A-3	โอเชียเนีย
7	Kamala Beach	idea, trip	0.83	0.88	1.02	A-3	อเมริกา
8	Kamala Beach	check, trip	0.83	0.88	1.01	A-3	โอเชียเนีย
9	Kamala Beach	check, trip	0.83	0.88	1.01	A-3	อเมริกา

เช่น การใช้คำสำคัญที่ได้จากการนับความถี่ของคำซึ่งจะส่งผลให้กฎความสัมพันธ์ที่ปรากฏเป็นไปตามคำสำคัญ แต่งานวิจัยนี้นำเอาการวัดความคล้ายคลึงกันของคำมาจัดกลุ่มคำตามโดเมนด้านการท่องเที่ยวทำให้สามารถลดกฎความสัมพันธ์ของคำที่มีความคล้ายคลึงกันเชิงความหมายได้

5. สรุปและอภิปรายผล

งานวิจัยนี้เสนอแบบจำลองการทำเหมืองข้อมูลกระดานสนทนาออนไลน์ด้านการท่องเที่ยวโดยใช้กฎความสัมพันธ์โดยค้นหารากศัพท์ของคำโดยใช้อัลกอริทึมของ Porter หาความคล้ายคลึงกันของคำเชิงความหมายโดยใช้อัลกอริทึมของ Wu and Palmer และหากฎความสัมพันธ์โดยใช้อัลกอริทึม FP-Growth ผลการทดลองจากวิจัยนี้เป็นกฎความสัมพันธ์ของนักท่องเที่ยวทั้ง 4 กลุ่มประเทศตามภูมิศาสตร์ นอกจากนี้งานวิจัยนี้ได้เปรียบเทียบกฎความสัมพันธ์ระหว่างสถานที่ท่องเที่ยวและคำ และกฎความสัมพันธ์ระหว่างสถานที่ท่องเที่ยวและโดเมนด้านการท่องเที่ยว ผลการทดลองสามารถสรุปตามการทดลองดังนี้

5.1 สรุปผลการทดลอง A: กฎความสัมพันธ์ระหว่างสถานที่ท่องเที่ยวและคำ

1) การทดลอง A-3 เป็นการทดลองที่เหมาะสมสำหรับค้นหากฎความสัมพันธ์ระหว่างกระตู่และความคิดเห็นจากคำในข้อความบนกระดานสนทนาออนไลน์ด้านการท่องเที่ยว เนื่องจากการเปรียบเทียบกฎความสัมพันธ์ระหว่างการทดลอง A-2 และ A-3 โดยใช้แหล่งข้อมูลจากความคิดเห็นแต่มีวิธีการแปลงข้อความที่ต่างกัน พบว่าการทดลอง A-3 เท่านั้นที่สามารถค้นหากฎความสัมพันธ์ที่มีค่าสนับสนุนและค่าความเชื่อมั่นในช่วงร้อยละ 80-100 ซึ่งเป็นช่วงค่าสนับสนุนและค่าความเชื่อมั่นสูงสุดของการทดลอง A

2) กฎความสัมพันธ์จากการทดลอง A สามารถแสดงความสัมพันธ์ระหว่างสถานที่และคำเท่านั้นแต่ยังไม่สามารถสรุปได้ว่ากฎความสัมพันธ์ที่ปรากฏนั้นถูกถามและแสดงความคิดเห็นในประเด็นหรือโดเมนใดเนื่องจากคำเหล่านี้เป็นคำที่ยังไม่ผ่านการจัดกลุ่มคำ

5.2 สรุปผลการทดลอง B: กฎความสัมพันธ์ระหว่างสถานที่ท่องเที่ยวและโดเมนด้านการท่องเที่ยว

1) แห่ลมพินา หาดกมลา พระพุทธรูปมิ่งมงคลเอก-นาคคีรี และหาดในหานถูกแสดงความคิดเห็นพร้อมกับโดเมนด้านการท่องเที่ยวสูงซึ่งผลการทดลอง B มีความสอดคล้องกับการทดลอง A คือ การทดลอง A พบว่าหาดกมลาเป็นสถานที่ท่องเที่ยวที่ถูกแสดงความคิดเห็นร่วมกับคำสูงดังตารางที่ 4 ในขณะที่เดียวกันการทดลอง B ดังตารางที่ 5 พบว่าหาดกมลาถูกแสดงความคิดเห็นพร้อมกับโดเมนด้านการท่องเที่ยวสูงเช่นกันนอกจากนี้การทดลอง B ยังพบอีกว่าแห่ลมพินาและหาดในหานถูกกล่าวถึงพร้อมกับโดเมนด้านการท่องเที่ยวสูง

2) โดเมน “Accommodation” “Time” และ “Contact” ถูกกล่าวถึงบนกระดานสนทนาออนไลน์สูงดังตารางที่ 5 นอกจากนี้โดเมนด้านการท่องเที่ยวทั้ง 3 โดเมนถูกกล่าวถึงร่วมกับสถานที่ท่องเที่ยวในจังหวัดภูเก็ตสูงเช่นกันเนื่องจากคำแต่ละโดเมนมีความสัมพันธ์กับสถานที่ท่องเที่ยว ตัวอย่างคำที่อยู่ในโดเมน “Accommodation” เช่น คำว่า “hotel” และ “resort” เป็นคำที่แสดงถึงที่พักในบริเวณสถานที่ท่องเที่ยว ส่วนคำที่อยู่ในโดเมน “Time” เช่น คำว่า “day” และ “night” เป็นคำที่แสดงถึงช่วงเวลาในการเดินทางหรือพักผ่อนในสถานที่ และคำที่อยู่ในโดเมน “Contact” เช่น คำว่า “taxi” และ “boat” เป็นคำที่แสดงถึงการติดต่อยานพาหนะเพื่อการเดินทางไปยังสถานที่ต่าง ๆ

3) การทดลอง B-3 เป็นการทดลองที่เหมาะสมสำหรับค้นหากฎความสัมพันธ์ระหว่างกระตู่และความคิดเห็นจากโดเมนด้านการท่องเที่ยวเนื่องจากผลการทดลอง B-2 และการทดลอง B-3 ที่ใช้แหล่งข้อมูลจากความคิดเห็นแต่มีวิธีการแปลงข้อความที่ต่างกันพบว่า กฎความสัมพันธ์โดยส่วนใหญ่ของการทดลอง B-3 มีค่าสนับสนุนสูงกว่าการทดลอง B-2

5.3 ผลการเปรียบเทียบระหว่างการทดลอง A และการทดลอง B

ชุดข้อมูลโดเมนด้านการท่องเที่ยวในการทดลอง B เหมาะสำหรับการนำไปใช้หากฎความสัมพันธ์ร่วมกับชุดข้อมูลสถานที่ท่องเที่ยวมากกว่าชุดข้อมูลคำของการทดลอง A เนื่องจากสามารถค้นพบกฎความสัมพันธ์

ของสถานที่ท่องเที่ยวในช่วงค่าสนับสนุนและค่าความเชื่อมั่นระหว่างร้อยละ 80-100 จำนวนสูงสุดถึง 13 สถานที่ดังแสดงในตารางที่ 5 และกฎความสัมพันธ์ที่ได้จากการทดลอง B ไม่เกิดกฎความสัมพันธ์ที่มีความหมายคล้ายคลึงกันระหว่างกฎเนื่องจากใช้โดเมนด้านการท่องเที่ยวเป็นชุดข้อมูลแสดงตัวอย่างในตารางที่ 7

จากผลการวิจัยข้างต้นทำให้ทราบถึงความสัมพันธ์ของสถานที่ท่องเที่ยว โดเมนด้านการท่องเที่ยว และค่าที่นักท่องเที่ยวกล่าวถึงบนกระดานสนทนาออนไลน์ด้านการท่องเที่ยวของจังหวัดภูเก็ตจะทราบการกล่าวถึงของนักท่องเที่ยวต่อสถานที่ท่องเที่ยวและโดเมนด้านการท่องเที่ยวเพื่อนำไปพัฒนาในด้านต่าง ๆ เช่น การคมนาคมความปลอดภัย และการบำรุงรักษาทัศนียภาพของสถานที่ท่องเที่ยวที่นักท่องเที่ยวสนใจ และหน่วยงานภาคเอกชนทราบการกล่าวถึงของนักท่องเที่ยวเพื่อนำไปใช้วางแผนการตลาดให้ตรงกับความต้องการของนักท่องเที่ยว อย่างไรก็ตามงานวิจัยนี้ยังมีข้อจำกัดในการเปรียบเทียบค่าความคล้ายคลึงกันของคำโดยเลือกจากคำที่มากที่สุดเพราะยังไม่สามารถจัดประเภทของคำตาม

ตารางที่ 6 จำนวนกฎความสัมพันธ์ที่มีค่าสนับสนุนและค่าความเชื่อมั่นระหว่างร้อยละ 80-100

ที่	การทดลอง	จำนวนสถานที่ท่องเที่ยว	
		การทดลอง A	การทดลอง B
1	ภายในกระทู้	3	13
2	กระทู้กับความเห็นทั้งหมด	3	3
3	กระทู้กับแต่ละความคิดเห็น	4	7

ตารางที่ 7 เปรียบเทียบกฎความสัมพันธ์

ที่	สถานที่ท่องเที่ยว	การทดลอง	
		การทดลอง A (คำ)	การทดลอง B (โดเมน)
1	Patong Beach	price	-
2	Patong Beach	rent	-
3	Patong Beach	-	Payment

ตารางที่ 8 เปรียบเทียบงานวิจัยนี้กับงานวิจัยที่เกี่ยวข้อง

ลำดับที่	บทความ	แหล่งข้อมูล	ชุดข้อมูล	ลบคำหยุด	ค้นหารากศัพท์	การวัดความคล้ายคลึงกันของคำ	อัลกอริทึมการหากฎความสัมพันธ์	ค่าสนับสนุนขั้นต่ำ	ค่าความเชื่อมั่นขั้นต่ำ
1	[22]	เว็บเชิงความหมาย	ชื่อคณะ และชื่ออาจารย์	ไม่ใช่	ไม่ใช่	ไม่ใช่	Apriori	3	ไม่ระบุ
2	[23]	เว็บไซต์	สถานที่ท่องเที่ยว และข้อมูลส่วนตัว	ไม่ใช่	ไม่ใช่	ไม่ใช่	Apriori	ไม่ระบุ	ไม่ระบุ
3	[24]	หนังสือพิมพ์	คำสำคัญ	ใช่	ใช่	ไม่ใช่	GARW	15	25
4	[25]	บทความ	คำสำคัญ	ใช่	ใช่	ไม่ใช่	ปรับปรุง Apriori	13	50
5	[26]	ข้อมูลทางการแพทย์	อาการ โรค ยา และอายุ	ใช่	ใช่	ไม่ใช่	FP-Growth และ Apriori	10	10
6	งานวิจัยนี้	กระดานสนทนาออนไลน์	สถานที่ท่องเที่ยว คำ และโดเมน	ใช่	ใช่	ใช่	FP-Growth	20	50

โดเมนได้ทั้งหมดเนื่องจากงานวิจัยนี้มีโดเมนด้านการท่องเที่ยวจำนวน 8 โดเมนเท่านั้นทำให้บางคำมีความคล้ายคลึงกันเท่ากันจึงไม่สามารถจัดประเภทของคำตามโดเมนได้ซึ่งส่วนนี้สามารถนำมาทดลองสำหรับงานวิจัยในอนาคต

6. เอกสารอ้างอิง

- [1]. Icoz, O., Kutuk, A., Icoz, O., "Social media and consumer buying decisions in tourism: the case of Turkey", *PASOS: Revista de Turismoy Patrimonio Cultural*, Vol. 16, No. 4, 1051-1066, 2018,
- [2]. Cong, G., Wang, L., Lin, C. Y., Song, Y. I., and Sun, Y. "Finding question-answer pairs from online forums.", *Proceeding of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, Singapore*, Singapore: 20 – 24 July, 2008.
- [3]. comScore, *comScore Ranks the Top 50 U.S. Digital Media Properties for June 2017*. Available Online at: <http://ir.comscore.com>, accessed on 18 August 2017.
- [4]. World Tourism Organization, *Tourism Satellite Account*. Available Online at: <http://statistics.unwto.org>, accessed on 15 June 2019.
- [5]. Ministry of Tourism and Sports of Thailand, *Tourism Statistics*, Available Online at: <http://www.mots.go.th>, accessed on 15 June 2019.
- [6]. Tourism Authority of Thailand PHUKET Guide Book, *Tourism Authority of Thailand*, Phuket Office, Thailand, n.d.
- [7]. Khruahong, S., Kong, X., and Hoang, D. "Ontology Design for Thailand Travel Industry". *International Journal of Knowledge Engineering*, Vol. 1, No. 3, pp. 191-196, 2015.
- [8]. Chochiang, K., Hanna, F., Betbeder, M. L., and Lapayre, J. C. "A new web-services based platform for semantic transliterated words searching to facilitate tourist strips in Thailand.", *Proceeding of Computer Supported Cooperative Work in Design*, Nanchang, China: 4-6 May, 2016.
- [9]. Namahoot, C. S., Brückner, M., and Panawong, N. "Context-Aware Tourism Recommender System Using Temporal Ontology and Naïve Bayes." *In Recent Advances in Information and Communication Technology*, pp. 183-194, 2015.
- [10]. Imsombut, A., and Sirikayon, C. "An alternative technique for populating Thai tourism ontology from texts based on machine learning." *Proceeding of Computer and Information Science (ICIS)*, Okayama, Japan: 26-29 June 2016
- [11]. Chowdhury, G. G. "Natural language processing." *Annual review of information science and technology*, Vol. 37, No. 1, pp. 51-89, 2003.
- [12]. Miller, G. A. "WordNet: a lexical database for English." *Communications of the ACM*, Vol. 38, No. 11, pp. 39-41, 1995.
- [13]. Liu, H., and Wang, P. "Assessing Sentence Similarity Using WordNet based Word Similarity." *JSW*, Vol. 8, No. 6, pp. 1451-1458, 2013.
- [14]. Wu, Z., and Palmer, M. "Verbs semantics and lexical selection.", *Proceeding of the 32nd annual meeting on Association for Computational Linguistics*, Las Cruces, New Mexico 27-30 June, 1994.
- [15]. Lo, R. T. W., He, B., and Ounis, I. "Automatically building a stopword list for an information retrieval system". *In Journal on Digital Information Management*, Vol. 5, pp.17-24, 2005.
- [16]. Frakes, W. B., and Baeza-Yates, R. Information retrieval: data structures and algorithms. *Prentice-Hall, Inc.*, U.S.A., 1992
- [17]. Alvares, R. V., Garcia, A. C. B., and Ferraz, I. "StEMBR: a stemming algorithm for the Brazilian Portuguese language." *In EPIA*, Vol. 3808, pp. 693-701, 2005.
- [18]. Porter, M. F. "An algorithm for suffix stripping." *Program*, Vol. 14, No. 3, pp. 130-137, 1980.



- [19] Han, J., Pei, J., and Kamber, M. Data mining: concepts and techniques, *Elsevier*, U.S.A., 2011.
- [20] Agrawal, R., and Srikant, R. "Fast algorithms for mining association rules." *In Proc. 20th int. conf. very large data bases*, VLDB, 1215, pp. 487-499, 1994.
- [21] Han, J., Pei, J., and Yin, Y. "Mining frequent patterns without candidate generation." *In ACM sigmod record*, 29(2), pp. 1-12, 2000.
- [22] Yazdi, A. S. H., and Kahani, M. "A novel model for mining association rules from semantic web data., *Proceeding of Intelligent Systems*, Bam, Iran: 4-6 February, 2014.
- [23] Aghdam, A. R., Kamalpour, M., Chen, D., Sim, A. T. H., and Hee, J. M. "Identifying places of interest for tourists using knowledge discovery techniques.", *Proceeding of Industrial Automation, Information and Communications Technology (IAICT)*, Bali, Indonesia: 28-30 August, 2014.
- [24] Kulkarni, M., and Kulkarni, S. "Knowledge Discovery in Text Mining using Association Rule Extraction." *International Journal of Computer Applications*, 143(12), 2016.
- [25] Mahgoub, H., and Rösner, D. "Mining association rules from unstructured documents." *World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering*, 2(8), 167-172, 2006
- [26] Lakshmi, K. S., and Kumar, G. S. "Association rule extraction from medical transcripts of diabetic patients.", *Proceeding of Applications of Digital Information and Web Technologies*, Bangalore, India: 17-19 February, 2014

