

ระบบสกัดความรู้จากบทวิจารณ์โรงแรมออนไลน์โดยใช้ตรรกศาสตร์คลุมเครือ

พิชญ์สินี กิจวัฒนาถาวร¹ ธรา อังสกุล² และ จิตมนต์ อังสกุล^{2*}

บทคัดย่อ

ปัจจุบัน วิธีการที่นักท่องเที่ยวนิยมนำมาใช้ในการเลือกสถานที่พักประกอบการท่องเที่ยวให้ได้ตรงกับความต้องการมากที่สุดคือ การวิเคราะห์ข้อมูลจากบทวิจารณ์ออนไลน์ของนักท่องเที่ยวที่เคยไปมาแล้วเพื่อนำมาใช้ประกอบการตัดสินใจ แต่ข้อเสียของการนำวิธีนี้มาใช้คือ นักท่องเที่ยวจะต้องอ่านบทวิจารณ์ทั้งหมดซึ่งมีเป็นจำนวนมาก บทความนี้จึงได้ออกแบบและพัฒนา ระบบสกัดความรู้แบบอัตโนมัติจากบทวิจารณ์โรงแรมออนไลน์ของนักท่องเที่ยว โดยมุ่งเน้นที่การวิเคราะห์ความหมายของบทวิจารณ์ของนักท่องเที่ยวเกี่ยวกับที่พัก โดยนำเอาออนโทโลยีมาใช้เป็นฐานความรู้ในการสกัดและ

จัดเก็บความรู้ และตัวแปลภาษามาช่วยในการวิเคราะห์เชิงความหมาย รวมทั้งพัฒนาวิธีการคำนวณระดับคะแนนความพึงพอใจของนักท่องเที่ยว โดยนำตรรกศาสตร์คลุมเครือมาใช้ ผลการทดลองพบว่า ระบบที่นำเสนอ มีค่าความถูกต้องของการสกัดความรู้เท่ากับ 90.15% ค่าความแม่นยำเท่ากับ 100% ค่าความระลึกเท่ากับ 75.12% และค่าระดับคะแนนของคำวิจารณ์โดยรวมแตกต่างจากที่นักท่องเที่ยวกำหนดเท่ากับ 0.06 คะแนน

คำสำคัญ: การสกัดความรู้ บทวิจารณ์ออนไลน์ ตรรกศาสตร์คลุมเครือ

¹ นักศึกษา สาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี

² อาจารย์ สาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 0-4422-4309 อีเมล: jitimon@sut.ac.th



A Knowledge Extraction System from Online Hotel Reviews Using Fuzzy Logic

Pichayasinee Kitwattanataworn¹ Thara Angskun² and Jitimon Angskun^{2*}

Abstract

Currently, the most common method that tourists use in selecting the accommodations to conform to their preferences is a data analysis from online reviews of experienced tourists to support the decision. Unfortunately, tourists must read enormous reviews in order to select their preferred accommodations. This paper designs and develops an automatic knowledge extraction system from online hotel reviews of tourists. The proposed system focuses on a semantic analysis from hotel reviews using ontology as the knowledge base

for knowledge extraction and storing, and the parser to solve the semantic analysis problem. In addition, we propose a calculation approach for a tourist satisfaction level using fuzzy logic. The experimental results indicate that the proposed system achieves with 90.15% of accuracy, 100% of precision, 75.12% of recall and 0.06 review points that are different from the points specified by tourists.

Keyword: Knowledge Extraction, Online Reviews, Fuzzy Logic

¹ Student, School of Information Technology, Institute of Social Technology, Suranaree University of Technology.

² Lecture, School of Information Technology, Institute of Social Technology, Suranaree University of Technology.

* Corresponding Author, Tel. 0-4422-4309, E-mail: jitimon@sut.ac.th

1. บทนำ

ในปัจจุบัน การพัฒนาทางด้านเทคโนโลยีสารสนเทศ และการสื่อสารมีความเจริญก้าวหน้าเป็นอย่างมาก ส่งผลให้อินเทอร์เน็ตได้เข้ามามีบทบาทในชีวิตประจำวันของผู้คนมากยิ่งขึ้น โดยทั่วโลกมีผู้ใช้งานอินเทอร์เน็ตสูงถึงประมาณ 1,966 ล้านคน (คิดเป็นร้อยละ 28.7 ของประชากรทั้งโลก) ในขณะที่ประเทศไทยมีผู้ใช้งานอินเทอร์เน็ตถึงประมาณ 17.5 ล้านคน [1] จากจำนวนการใช้งานอินเทอร์เน็ตที่สูงมากดังกล่าว ได้ส่งผลให้การดำเนินกิจกรรมทางธุรกิจแบบดั้งเดิมเปลี่ยนรูปแบบไปเป็นธุรกิจอิเล็กทรอนิกส์ (E-business) มากขึ้น

ธุรกิจการท่องเที่ยวนับเป็นธุรกิจหลักอย่างหนึ่งของประเทศไทย ซึ่งปัจจุบันการดำเนินการขององค์กรต่างๆ ที่มีส่วนเกี่ยวข้องกับธุรกิจนี้ได้มีการปรับเปลี่ยนองค์กรของตนเองไปสู่ธุรกิจอิเล็กทรอนิกส์มากยิ่งขึ้น เพื่อเป็นแรงจูงใจให้นักท่องเที่ยวมาซื้อสินค้าหรือบริการของตน และอำนวยความสะดวกแก่นักท่องเที่ยวในการบริการจองโรงแรมที่พัก จองรถ หรือจองตั๋วเครื่องบิน ตลอดจนช่วยให้นักท่องเที่ยวสามารถสืบค้นข้อมูลด้านการท่องเที่ยว เพื่อนำมาประกอบการตัดสินใจเลือกแหล่งท่องเที่ยว ได้ตรงกับความต้องการ ซึ่งแนวโน้มดังกล่าวได้สอดคล้องกับผลการศึกษาขององค์การส่งเสริมการท่องเที่ยวโลก (World Tourism Organization: WTO) ที่พบว่าปัจจัยสำคัญที่ส่งผลให้การท่องเที่ยวโลกเพิ่มสูงขึ้น คือ ระบบข้อมูลข่าวสารที่ทันสมัยและก้าวหน้า เนื่องจากเมื่อนักท่องเที่ยวมีโอกาสรับทราบข้อมูลข่าวสารด้านการท่องเที่ยวมากขึ้น จะทำให้มีความสนใจและมั่นใจในการเดินทางเพิ่มมากขึ้นตามไปด้วย [2]

อย่างไรก็ตาม ถึงแม้ว่าการจัดเก็บข้อมูลต่างๆ จากนักท่องเที่ยวจะมีการปฏิบัติกันอยู่แล้วอย่างกว้างขวางในโลกธุรกิจอิเล็กทรอนิกส์ แต่การจะนำข้อมูลเหล่านั้นมาใช้ประโยชน์กลับเป็นเรื่องที่ยุ้งยากและต้องใช้เวลา นาน เนื่องจากข้อมูลที่จัดเก็บไว้นั้นเป็นข้อมูลดิบที่ยังไม่ได้ผ่านการคัดกรอง ประมวลผล หรือสกัด เพื่อให้ได้ความรู้ที่สามารถนำไปใช้ประกอบการตัดสินใจได้ทันที ตัวอย่างเช่น เว็บไซต์สำหรับการท่องเที่ยวที่มีการจัดเก็บ

บทวิจารณ์เกี่ยวกับโรงแรมหรือที่พักของนักท่องเที่ยวที่เคยไปมาแล้ว เพื่อเป็นข้อมูลให้กับนักท่องเที่ยวคนอื่นๆ ซึ่งยังไม่มีความต้องการที่ชัดเจนว่าจะไปท่องเที่ยวที่ไหน และต้องการค้นหาข้อมูลเพื่อนำไปใช้ประกอบการวางแผนท่องเที่ยวของตน เว็บไซต์ประเภทนี้จะให้เพียงข้อมูลอันดับความนิยมของโรงแรมหรือที่พัก โดยรวมที่นักท่องเที่ยวคนอื่นได้เคยแสดงความคิดเห็นด้วยวิธีการลงคะแนนเท่านั้น หากนักท่องเที่ยวคนใดต้องการรายละเอียดเพิ่มเติมจะต้องตามอ่านในทุกๆ ความคิดเห็นหรือทุกๆ บทวิจารณ์เอง แล้วจึงวิเคราะห์หรือตัดสินใจว่าโรงแรมหรือที่พักนั้นควรไปหรือไม่ อีกทั้งเนื่องจากเว็บไซต์เป็นแบบตายตัว ข้อมูลที่จัดเก็บไว้อย่างไม่ได้ผ่านการคัดกรอง ทำให้ข้อมูลที่ได้อาจยังไม่ละเอียดเพียงพอสำหรับความต้องการของนักท่องเที่ยว

จากการทบทวนวรรณกรรมที่ศึกษาเกี่ยวกับบทวิจารณ์ออนไลน์พบว่า งานวิจัยส่วนใหญ่เป็นการจำแนกบทวิจารณ์ให้อยู่ในหมวดหมู่ที่ต้องการแบบอัตโนมัติ ตัวอย่างเช่น การสร้างขั้นตอนวิธีอย่างง่ายในการแยกประเภทคำวิจารณ์ว่ามีความหมายเชิงแนะนำหรือไม่ โดยการพยากรณ์จากการคำนวณค่าเฉลี่ยของกลุ่มคำคุณศัพท์หรือคำกริยาวิเศษณ์ [3] การจำแนกข้อความแบบอัตโนมัติโดยนำวิธีมาตรฐานในการคำนวณความหมายจากตำแหน่งของคำคุณศัพท์ในข้อความมาประยุกต์ใช้ [4] และการวิเคราะห์บทวิจารณ์ทางด้านดนตรี และพัฒนาระบบที่มีกลไกในการกำจัดอคติตลอดจนการบรรยายที่ไม่เกี่ยวข้องกับดนตรีของผู้วิจารณ์ โดยการเรียนรู้จากบทวิจารณ์จำนวนมาก [5]

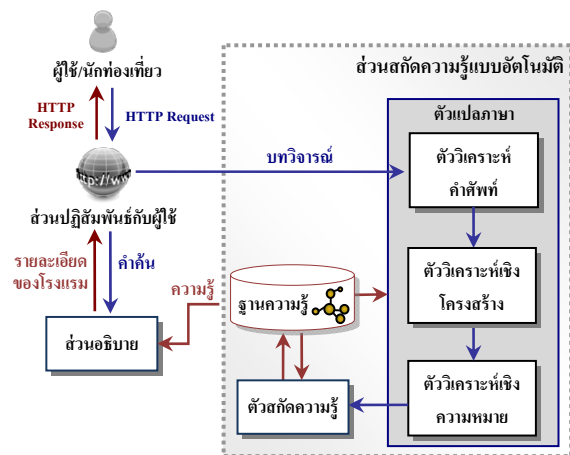
นอกจากนี้ยังมีงานวิจัยที่ศึกษาเกี่ยวกับการสกัดความรู้จากบทวิจารณ์ออนไลน์ ตัวอย่างเช่น การพัฒนาระบบจัดอันดับสินค้าจากคุณลักษณะที่ปรากฏในบทวิจารณ์ โดยมีกระบวนการระบุประเภทและแปลความหมายของประโยค รวมทั้งการสร้างกราฟเพื่อเทียบเคียงคุณภาพของสินค้าต่างรุ่นหรือต่างผู้ผลิต แล้วนำไปคำนวณระดับคะแนนเพื่อจัดอันดับ [6] การพัฒนาระบบสกัดความรู้จากบทวิจารณ์สินค้า โดยมุ่งเน้นการค้นหาคุณลักษณะของสินค้าและนำคุณลักษณะที่ได้มาใช้คำนวณหาระดับ

คะแนน อีกทั้งมีการนำเทคนิคตรรกศาสตร์คลุมเครือมาใช้ในการจำแนกและคำนวณหาระดับความเป็นข้อความขยะในแต่ละบทวิจารณ์ [7] และการสกัดความรู้จากบทวิจารณ์โดยการสรุปสาระสำคัญและนำเสนอทั้งในด้านบวกและลบให้กับลูกค้า ด้วยวิธีการทำเหมืองข้อมูลและการประมวลผลภาษาธรรมชาติ [8]

จากการทบทวนวรรณกรรมข้างต้นจะเห็นได้ว่าลูกค้าไม่สามารถทราบถึงจำนวนคำวิจารณ์ต่างๆ อย่างละเอียดว่ามีผู้ที่แสดงความคิดเห็นทั้งด้านบวกและด้านลบในแต่ละเรื่องจำนวนเท่าใด ซึ่งในการจัดลำดับความสำคัญนั้นจำนวนของการแสดงความคิดเห็นนับว่ามีผลอย่างมากต่อการจัดลำดับความสำคัญหรือการตัดสินใจเลือกสินค้าและบริการของลูกค้า [9] ดังนั้น งานวิจัยนี้จึงนำเสนอระบบสกัดความรู้จากบทวิจารณ์ออนไลน์ของนักท่องเที่ยวซึ่งมีการแสดงจำนวนความคิดเห็นของแต่ละเรื่องอย่างละเอียด และนำเอาเทคนิคต่างๆ ที่มีอยู่ในปัจจุบันเข้ามาประยุกต์ใช้ เช่น ออนโทโลยี มาใช้เป็นฐานความรู้ในการวิเคราะห์คำศัพท์ การสกัดและจัดเก็บความรู้ นอกจากนี้ยังนำตัวแปลภาษามาช่วยในการวิเคราะห์เชิงโครงสร้างและเชิงความหมาย รวมทั้งพัฒนาวิธีการคำนวณระดับคะแนนความพึงพอใจโดยนำตรรกศาสตร์คลุมเครือมาใช้ เพื่อให้ได้มาซึ่งผลลัพธ์ที่มีความใกล้เคียงกับความพึงพอใจของนักท่องเที่ยวมากที่สุด

2. กรอบการทำงานของระบบ

ในการพัฒนาระบบสกัดความรู้แบบอัตโนมัติจากบทวิจารณ์โรงแรมออนไลน์ มุ่งเน้นการพัฒนาขั้นตอนวิธีในการประมวลผลความหมายอย่างถูกต้องจากบทวิจารณ์ภาษาอังกฤษของนักท่องเที่ยวที่ได้ไปแหล่งท่องเที่ยวมาแล้ว โดยการสกัดเอาส่วนสำคัญของบทวิจารณ์มาเป็นความรู้ที่เก็บอยู่ในฐานความรู้ของระบบ และเมื่อนักท่องเที่ยวคนใหม่เข้ามาค้นหาข้อมูล ระบบจะสืบค้นความรู้ที่เก็บอยู่ในฐานความรู้นั้นไปแสดงผล โดยกรอบการทำงานของระบบสกัดความรู้จากบทวิจารณ์โรงแรมออนไลน์ แสดงดังรูปที่ 1 อันประกอบด้วย 5 ส่วน ได้แก่ ส่วนปฏิสัมพันธ์กับผู้ใช้ ฐานความรู้ ตัวแปลภาษา ตัวสกัดความรู้ และ



รูปที่ 1 กรอบการทำงานของระบบสกัดความรู้จากบทวิจารณ์โรงแรมออนไลน์

ส่วนอธิบายความรู้ โดยสามารถอธิบายการทำงานในแต่ละส่วนดังต่อไปนี้ [10]

2.1 ส่วนปฏิสัมพันธ์กับผู้ใช้

เป็นส่วนโต้ตอบกับผู้ใช้หรือนักท่องเที่ยวซึ่งใช้ระบบผ่านทางเว็บเบราว์เซอร์

2.2 ฐานความรู้

เป็นส่วนที่ใช้เก็บข้อมูลที่เป็นประโยชน์ต่อการสกัดความรู้แบบอัตโนมัติ ตลอดจนความรู้ที่สกัดได้ ซึ่งประกอบด้วย 4 ส่วนหลัก ดังนี้

2.2.1 ออนโทโลยีด้านการท่องเที่ยว ได้ปรับปรุงจากออนโทโลยีที่มีอยู่ในปัจจุบัน [11] โดยการนำบทวิจารณ์ของนักท่องเที่ยวจำนวน 400 บทวิจารณ์มาวิเคราะห์หาคำสำคัญ (Feature Selection) ที่อยู่ในขอบเขตเรื่องโรงแรมหรือที่พัก โดยใช้วิธีการวัดค่าน้ำหนักของคำตามขั้นตอนวิธีของร็อคชิโอ (Rocchio: TF-IDF Weighting) [12] จากนั้นจึงนำคำสำคัญที่ได้มาเพิ่มเติมลงในออนโทโลยีดังกล่าว ทั้งนี้โครงสร้างของออนโทโลยีด้านการท่องเที่ยวที่ปรับปรุงแล้วประกอบด้วยคลาส (Class) จำนวน 10 คลาส และมีคุณสมบัติ (Properties) ที่สำคัญ

ต่อการสกัดความรู้รวมทั้งสิ้น 76 คุณสมบัติ ดังแสดง ตัวอย่างของออนโทโลยีบางส่วนไว้ในรูปที่ 2

จากตัวอย่างโครงสร้างของออนโทโลยีด้านการท่องเที่ยว ในรูปที่ 2 จะเห็นได้ว่าคลาสโรงแรม ประกอบด้วยคลาสนย่อย คือ ห้องพัก ห้องน้ำ ทำเลที่ตั้ง การบริการ และความ คุ่มค่า และในแต่ละคลาสประกอบด้วยคุณสมบัติต่างๆ เช่น คลาสห้องพัก ประกอบด้วยคุณสมบัติคือ เครื่อง ปรับอากาศ ผ้าปูเตียง และเครื่องโทรทัศน์ และคลาส การจัดอันดับ ประกอบด้วยคุณสมบัติคือ ดีมาก ดี ปานกลาง แย่ และแย่มาก โดยคลาสการจัดอันดับนี้จะ ถูกนำมาใช้ในการจัดเก็บระดับคะแนนความพึงพอใจของ นักท่องเที่ยวเกี่ยวกับโรงแรม ซึ่งคำนวณได้จากตัวสกัด ความรู้ที่จะกล่าวต่อไปในหัวข้อ 2.4

2.2.2 คลังคำเหมือน เป็นแหล่งรวบรวมคำต่างๆ ที่มีความหมายเหมือนกัน แต่เขียนต่างกัน โดยมีการกำหนด คำที่เป็นตัวแทนของกลุ่มคำเหมือนแต่ละกลุ่ม เพื่อใช้ในการแปลงคำศัพท์ให้เป็นคำเดียวกันทั้งหมด

2.2.3 คลังคำศัพท์ เป็นแหล่งรวบรวมคำศัพท์ที่มีการกำหนดระดับคะแนนของคำไว้โดยผู้เชี่ยวชาญ ด้านภาษา และคะแนนแต่ละระดับจะแสดงถึงความ พึงพอใจที่แตกต่างกัน ได้แก่ ระดับคะแนนที่ 1 หมายถึง แย่มาก (Terrible) ระดับคะแนนที่ 2 หมายถึง แย่ (Poor) ระดับคะแนนที่ 3 หมายถึง ปานกลาง (Average) ระดับ คะแนนที่ 4 หมายถึง ดี (Good) และระดับคะแนนที่ 5 หมายถึง ดีมาก (Excellent) คลังคำศัพท์ประกอบด้วย กลุ่มคำศัพท์ 3 ประเภทดังนี้

1) กลุ่มคำคุณศัพท์ (adj.) เป็นกลุ่มคำที่แสดง การติ-ชม ซึ่งมีการกำหนดระดับคะแนนของแต่ละคำ ไว้ตายตัวโดยการยืนยันจากผู้เชี่ยวชาญด้านภาษา เช่น “Excellent” = 5, “Effective” = 4, “Moderate” = 3 และ “Unfriendly” = 2, “Awful” = 1 เป็นต้น

2) กลุ่มคำกริยาวิเศษณ์ (adv.) เป็นกลุ่มคำแสดง การขยายความติ-ชมไม่ว่าจะเป็นแง่บวกหรือลบ ซึ่งมีการกำหนดอัตราการเพิ่ม-ลดระดับคะแนนไว้ ทั้งนี้ เมื่อนำคำกลุ่มนี้ไปประมวลผลจะทำให้ระดับคะแนนที่สกัดได้ มีผลลัพธ์ไปในทางเดียวกัน กล่าวคือ หากผู้ใช้วิจารณ์ใน

➤ Hotel (โรงแรม)

- Roo (ห้องพัก)
 - AirCondition (เครื่องปรับอากาศ)
 - Bedsheet (ผ้าปูเตียง)
 - TVSet (เครื่องโทรทัศน์)
 - ...
- Bathroo (ห้องน้ำ)
 - Bathtub (อ่างอาบน้ำ)
 - Toilet (ห้องสุขา)
 - ...
- Location (ทำเลที่ตั้ง)
- Service (การบริการ)
 - Employee (พนักงาน)
 - Reception (พนักงานต้อนรับ)
 - ...
- Value (ความคุ้มค่า)

➤ Rating (การจัดอันดับ)

- Excellent (ดีมาก)
- Good (ดี)
- Average (ปานกลาง)
- Poor (แย่)
- Terrible (แย่มาก)

รูปที่ 2 โครงสร้างของออนโทโลยีด้านการท่องเที่ยว

แง่ดี ระดับคะแนนที่ได้ก็จะยิ่งเพิ่มขึ้น ตัวอย่างกลุ่มคำกริยาวิเศษณ์ที่ได้กำหนดอัตราการเพิ่ม-ลดระดับคะแนนไว้ เช่น “Very” = ± 1 , “So” = ± 1 , “Extremely” = ± 1 และ “Most” = ± 2 เป็นต้น

3) กลุ่มคำกริยาวิเศษณ์แบบพิเศษ (sadv.) เป็นกลุ่มคำแสดงการขยายความติ-ชมทั้งแง่บวกและลบ ซึ่งมีการกำหนดอัตราการเพิ่ม-ลดระดับคะแนนไว้เช่นเดียวกับกลุ่มคำกริยาวิเศษณ์ แต่แตกต่างกันตรงที่เมื่อนำคำกลุ่มนี้ไปประมวลผลจะทำให้ระดับคะแนนที่สกัดได้มีผลลัพธ์ไปในทางตรงกันข้าม กล่าวคือ หากผู้ใช้วิจารณ์ในแง่ดี ระดับคะแนนที่สกัดได้ก็จะถูกลดทอนลง ตัวอย่างกลุ่มคำกริยาวิเศษณ์ที่ได้กำหนดอัตราการเพิ่ม-ลดระดับคะแนนไว้ เช่น “Most” = ± 1 , “So” = ± 1 , “Extremely” = ± 1 และ “Most” = ± 2 เป็นต้น

ทั้งนี้ ระดับคะแนนของคำศัพท์ทั้ง 3 ประเภทจะถูกนำไปใช้ในตัวสกัดความรู้ ซึ่งจะกล่าวต่อไปในหัวข้อ 2.4

2.2.4 กฎไวยากรณ์ไม่พึ่งบริบท (Context Free Grammar) เป็นกลุ่มของกฎไวยากรณ์ที่เขียนขึ้นโดยแสดงเป็นสัญลักษณ์ทางภาษา [13] เพื่อใช้สำหรับตรวจสอบไวยากรณ์ของประโยคคำวิจารณ์ในขั้นตอนการแปลภาษา

2.3 ตัวแปลภาษา

เป็นส่วนสำคัญของระบบที่ทำให้ทราบกลุ่มคำสำคัญที่ใช้ในการวิเคราะห์ความหมายทั้งหมด สามารถแบ่งการทำงานออกเป็น 3 ส่วน ดังต่อไปนี้

2.3.1 ตัววิเคราะห์คำศัพท์ (Lexical Analyzer) ทำหน้าที่ในการตัดคำตามช่องว่าง และแปลงคำศัพท์ให้อยู่ในรูปแบบที่เหมาะสมสำหรับการสกัดความรู้ โดยเปรียบเทียบกับคลังคำเหมือนที่อยู่ในฐานความรู้

2.3.2 ตัววิเคราะห์เชิงโครงสร้าง (Syntactic Analyzer) ทำหน้าที่วิเคราะห์หาความเกี่ยวข้องสัมพันธ์ระหว่างคำในประโยคตามทฤษฎีไวยากรณ์ไม่พึ่งบริบท

2.3.3 ตัววิเคราะห์เชิงความหมาย (Semantic Analyzer) ทำหน้าที่วิเคราะห์ความหมายของคำที่ได้จากตัววิเคราะห์เชิงโครงสร้าง โดยการค้นหาระดับคะแนนของคำคุณศัพท์และคำกริยาวิเศษณ์จากคลังคำศัพท์

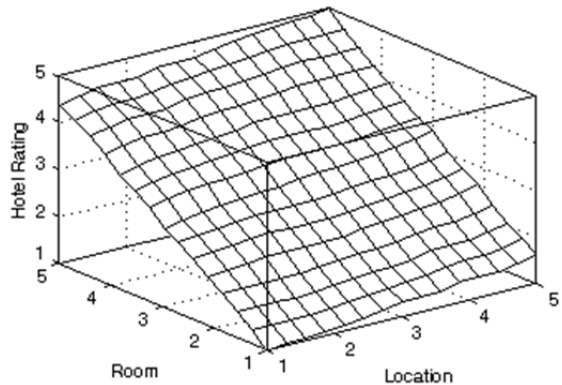
2.4 ตัวสกัดความรู้

เป็นการคำนวณระดับคะแนนตามคำคุณศัพท์และคำกริยาวิเศษณ์ที่ปรากฏในประโยคคำวิจารณ์ แบ่งการคำนวณออกเป็น 2 ระดับ ดังนี้

2.4.1 การคำนวณระดับคะแนนเฉพาะเรื่อง เป็นการคำนวณค่าคะแนนความพึงพอใจเฉพาะในเรื่องที่คำนาม (Noun) แสดงถึง โดยระบบจะพิจารณาค่าระดับคะแนนของกลุ่มคำคุณศัพท์ตามทีระบุไว้ในคลังคำศัพท์ เมื่อทราบแล้วจึงพิจารณาต่อไปว่ามีกลุ่มคำกริยาวิเศษณ์ที่เกี่ยวข้องหรือไม่ ถ้ามีจะต้องปรับระดับคะแนนของเรื่องนั้นใหม่ ทั้งนี้ ค่าของระดับคะแนนเฉพาะเรื่องที่คำนวณได้จะถูกจัดเก็บลงในออนโทโลยีโดยจำแนกตามเรื่องหรือคุณสมบัติของโรงแรมที่วิจารณ์ เพื่อร่อนนำไปแปลความหมายตามฟังก์ชัน



รูปที่ 3 ตัวอย่างการคำนวณระดับคะแนนเฉพาะเรื่อง



รูปที่ 4 ตัวอย่างการพิจารณาระดับคะแนนในขั้นก่อนหน้า เพื่อคำนวณหาระดับคะแนนในภาพรวมของทั้งโรงแรม

สมาชิกของตรรกศาสตร์คลุมเครือในขั้นตอนการคำนวณระดับคะแนนในภาพรวมต่อไป ตัวอย่างการคำนวณระดับคะแนนของเครื่องปรับอากาศ (Air Condition) ที่ได้จากประโยคคำวิจารณ์ “The air conditioning was almost very effective.” แสดงดังรูปที่ 3

2.4.2 การคำนวณระดับคะแนนในภาพรวม ระดับคะแนนเฉพาะเรื่องที่คำนวณได้จากขั้นตอนที่ผ่านมาจะถูกนำมาเป็นข้อมูลนำเข้าในระบบวินิจฉัยแบบฟัซซี่ (Fuzzy Inference System) เพื่อคำนวณระดับคะแนนในภาพรวม โดยระดับคะแนนในภาพรวมนั้นจะแบ่งเป็นลำดับขั้นตามโครงสร้างของออนโทโลยีดังที่กล่าวไปแล้ว โดยระดับคะแนนในภาพรวมของลำดับขั้นล่างจะถูกนำไปใช้เป็นข้อมูลนำเข้าในระบบวินิจฉัยแบบฟัซซี่ในลำดับขั้นบนถัดขึ้นไปเรื่อยๆ จนกลายเป็นระดับคะแนนในภาพรวมของทั้งโรงแรม ดังตัวอย่างที่แสดงในรูปที่ 4 ซึ่งเป็น การหาระดับคะแนนในภาพรวมของทั้งโรงแรม (Hotel Rating)

โดยพิจารณาจากระดับคะแนนในภาพรวมของชั้นก่อนหน้า เช่น ห้องพัก (Room) และทำเลที่ตั้ง (Location) เป็นต้น

- ทั้งนี้ ระบบวินิจฉัยแบบฟัซซีที่นำมาใช้ในการคำนวณระดับคะแนนในภาพรวมนั้น มีส่วนประกอบพื้นฐานที่สำคัญอยู่ 4 ส่วน ดังนี้

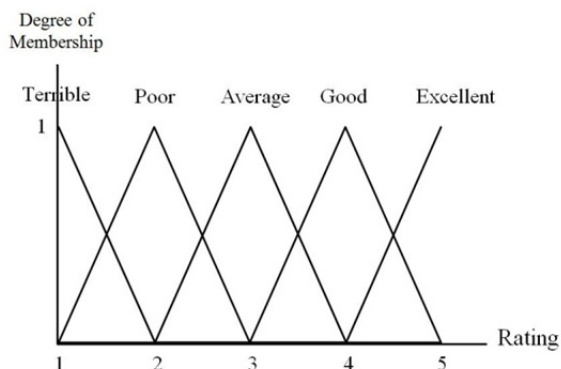
- **ส่วนฟัซซิฟิเคชัน (Fuzzification)** เป็นการนำเอาค่าของระดับคะแนนเฉพาะเรื่องมาแปลงให้เป็นค่าแบบฟัซซี (Fuzzy Values) ซึ่งกระบวนการนี้ต้องมีการกำหนดฟัซซีเซตสำหรับคุณสมบัติของโรงแรมที่มีอยู่ในออนไลน์ไทยทั้ง 76 คุณสมบัติ ซึ่งฟัซซีเซตของแต่ละคุณสมบัติได้กำหนดรูปแบบของระดับค่าความเป็นสมาชิก (Degree of Membership) แบบฟังก์ชันสามเหลี่ยม (Triangular Membership Function) ซึ่งประกอบด้วยพารามิเตอร์ 3 ค่า ได้แก่ $\{a, b, c\}$ ดังสมการที่ 1 [14]

$$\text{Triangular}(x; a, b, c) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \leq x \leq b \\ (c-x)/(c-b) & b \leq x \leq c \\ 0 & x > c \end{cases} \quad (1)$$

ฟังก์ชันสามเหลี่ยมนี้ได้ถูกนำมาใช้แบ่งกลุ่มสมาชิกของตัวแปรระดับคะแนน (Rating) ออกเป็น 5 กลุ่มสมาชิก ได้แก่ แย่มาก (Terrible) แย่ (Poor) ปานกลาง (Average) ดี (Good) และดีมาก (Excellent) โดยกำหนดค่าพารามิเตอร์ในแต่ละฟังก์ชันสมาชิกแตกต่างกัน อาทิ การแบ่งกลุ่มสมาชิกตัวแปรที่มีคะแนนอยู่ในระดับแย่ (Poor) กำหนดค่าพารามิเตอร์ให้ $a=1, b=2$ และ $c=3$ เป็นต้น ทั้งนี้ การแบ่งกลุ่มสมาชิกของตัวแปรระดับคะแนน แสดงดังรูปที่ 5

- **ส่วนฐานกฎฟัซซี (Fuzzy Rule Base)** ประกอบด้วยกฎฟัซซีต่างๆ ที่ใช้ในการอนุมานจากเหตุไปสู่ผลลัพธ์ นั่นก็คือ หากนักท่องเที่ยววิจารณ์เกี่ยวกับคุณสมบัติต่างๆ ของโรงแรม จะอนุมานหรือสรุประดับคะแนนในภาพรวมของโรงแรมได้ ซึ่งนิยมเขียนแสดงความสัมพันธ์อยู่ในรูปของ ถ้า-แล้ว (If-Then) อาทิ “If Location=Good Then Hotel=Good”

ในการคำนวณระดับคะแนนในภาพรวมโดยใช้



รูปที่ 5 การแบ่งกลุ่มสมาชิกตัวแปรระดับคะแนน

ตรรกศาสตร์คลุมเครือนี้ ผู้วิจัยได้สร้างกฎขึ้นมาจำนวนทั้งสิ้น 395 กฎ แบ่งเป็นกฎของฟัซซีเซตเรื่องห้องน้ำ 40 กฎ ห้องพัก 135 กฎ การบริการ 25 กฎ และภาพรวมของทั้งโรงแรม 195 กฎ ซึ่งกฎของฟัซซีเซตแต่ละเรื่องนี้จะมีการให้ค่าน้ำหนักของแต่ละกฎที่แตกต่างกันไป โดยการกำหนดน้ำหนักของแต่ละกฎจะเป็นไปตามแนวคิดที่ว่า “คุณสมบัติของโรงแรมที่กำหนดในแต่ละกฎนั้น สอดคล้องกับระดับคะแนนในภาพรวมของทั้งโรงแรมมากน้อยเพียงไร” ตัวอย่างเช่น เมื่อนักท่องเที่ยววิจารณ์ว่าโรงแรมนี้มีทำเลที่ตั้งดี (คิดเป็นระดับ 4 คะแนน) แต่คะแนนในภาพรวมของบทวิจารณ์ถูกกำหนดว่าโรงแรมนี้ได้เพียงระดับที่ 1 คือแย่มาก แสดงว่าคุณสมบัติเรื่องทำเลที่ตั้งสอดคล้องกับระดับคะแนนในภาพรวมของทั้งโรงแรมน้อยมาก ค่าน้ำหนักของกฎเรื่องทำเลที่ตั้งจึงควรกำหนดให้เหมาะสม ในที่นี้ค่าน้ำหนักของกฎจะกำหนดให้มีค่า 3 ค่า คือ 0.1, 0.5 และ 1 ซึ่งค่าน้ำหนักมาก (เท่ากับ 1) หมายถึง การวิจารณ์ในเรื่องนั้นสอดคล้องกับระดับคะแนนในภาพรวมมาก ค่าน้ำหนักปานกลาง (เท่ากับ 0.5) หมายถึง การวิจารณ์ในเรื่องนั้นสอดคล้องกับระดับคะแนนในภาพรวมปานกลาง และค่าน้ำหนักน้อย (เท่ากับ 0.1) หมายถึง การวิจารณ์ในเรื่องนั้นสอดคล้องกับระดับคะแนนในภาพรวมน้อย

โดยการพิจารณาว่าคุณสมบัติต่างๆ สอดคล้องกับระดับคะแนนในภาพรวมของทั้งโรงแรมมากน้อยเพียงไร จะได้มาจากการวิเคราะห์การให้คะแนนของกลุ่มตัวอย่าง

ซึ่งก็คือบทวิจารณ์ที่สุ่มเก็บมาจากเว็บไซต์ TripAdvisor.com จำนวน 100 บทวิจารณ์ โดยมีระดับความพึงพอใจในภาพรวมของทั้งโรงแรมตามที่นักท่องเที่ยวให้คะแนนแตกต่างกัน 5 ระดับคะแนน จำนวนระดับคะแนนและ 20 บทวิจารณ์ โดยพิจารณาว่า คะแนนเฉลี่ยในเรื่องต่างๆ ที่นักท่องเที่ยวกำหนดให้ในแต่ละระดับมีความใกล้เคียงกับคะแนนในภาพรวมของทั้งโรงแรมกี่ระดับ

ในตารางที่ 1 เป็นข้อมูลการให้คะแนนของนักท่องเที่ยวจากกลุ่มตัวอย่างบทวิจารณ์ ในเรื่องทำเลที่ตั้ง พบว่ามี 3 บทวิจารณ์ที่กล่าวถึงทำเลที่ตั้งของโรงแรมระดับที่ 1 โดยนักท่องเที่ยวให้คะแนนในภาพรวมของทั้งโรงแรมแย่มาก (ระดับที่ 1) แต่วิจารณ์ว่าโรงแรมนั้นมีทำเลที่ตั้งดี (โดยเฉลี่ย 3 บทวิจารณ์ได้ระดับคะแนนเฉพาะเรื่องทำเลที่ตั้ง 4 คะแนน) เป็นต้น

ตารางที่ 1 ตัวอย่างการให้คะแนนของนักท่องเที่ยวในเรื่องทำเลที่ตั้ง จากกลุ่มตัวอย่างบทวิจารณ์

ระดับคะแนน ในภาพรวมของทั้งโรงแรม	ระดับคะแนน เรื่องทำเลที่ตั้ง
ระดับที่ 1 ★	พบการวิจารณ์ 3 ครั้ง (คะแนนเฉลี่ย 4 คะแนน)
ระดับที่ 2 ★★	พบการวิจารณ์ 6 ครั้ง (คะแนนเฉลี่ย 4 คะแนน)
ระดับที่ 3 ★★★	พบการวิจารณ์ 10 ครั้ง (คะแนนเฉลี่ย 4.4 คะแนน)
ระดับที่ 4 ★★★★	พบการวิจารณ์ 13 ครั้ง (คะแนนเฉลี่ย 4.31 คะแนน)
ระดับที่ 5 ★★★★★	พบการวิจารณ์ 7 ครั้ง (คะแนนเฉลี่ย 4.57 คะแนน)

จากตารางที่ 1 แสดงให้เห็นว่า ผู้ใช้ให้คะแนนในเรื่องทำเลที่ตั้งมีความใกล้เคียงกับคะแนนในภาพรวมของทั้งโรงแรมจำนวน 2 ระดับ (คือ ระดับ 4 และระดับ 5) แต่ในระดับ 1, 2 และ 3 มีความแตกต่างกับการวิจารณ์เรื่องทำเลที่ตั้ง ตามเกณฑ์ความใกล้เคียงที่กำหนดไว้ในตารางที่ 2

ตารางที่ 2 เกณฑ์การพิจารณาความใกล้เคียงระหว่างระดับคะแนนในเรื่องต่างๆ กับระดับคะแนนในภาพรวมของทั้งโรงแรม

เกณฑ์ความใกล้เคียง
คะแนนเฉลี่ย 1.00-1.99 กำหนดค่าให้ใกล้กับระดับที่ 1
คะแนนเฉลี่ย 1.01-2.99 กำหนดค่าให้ใกล้กับระดับที่ 2
คะแนนเฉลี่ย 2.01-3.99 กำหนดค่าให้ใกล้กับระดับที่ 3
คะแนนเฉลี่ย 3.01-4.99 กำหนดค่าให้ใกล้กับระดับที่ 4
คะแนนเฉลี่ย 4.01-5.00 กำหนดค่าให้ใกล้กับระดับที่ 5

จากการพิจารณาว่า คะแนนเฉลี่ยในเรื่องต่างๆ มีความใกล้เคียงกับคะแนนในภาพรวมของทั้งโรงแรมกี่ระดับ จะนำค่าจำนวนเป็นร้อยละความน่าจะเป็นของการวิจารณ์เรื่องนั้นๆ ว่าสอดคล้องกับระดับคะแนนในภาพรวมของทั้งโรงแรมเท่าใด ซึ่งจากการวิเคราะห์เรื่องทำเลที่ตั้งจะได้ว่ามีค่าความน่าจะเป็นเพียง 40% ($2/5 \times 100$) เนื่องจากมีความใกล้เคียงเพียง 2 ระดับจาก 5 ระดับ

จากนั้นจึงนำค่าร้อยละความน่าจะเป็นดังกล่าวมาแปลงเป็นค่าน้ำหนักของกฎตามเกณฑ์ที่แสดงในตารางที่ 3

ตารางที่ 3 เกณฑ์การพิจารณาค่าน้ำหนักจากร้อยละความน่าจะเป็นที่การวิจารณ์เฉพาะเรื่องหนึ่งๆ จะส่งผลต่อระดับคะแนนในภาพรวม

ร้อยละความน่าจะเป็น	ค่าน้ำหนักของกฎ
มากกว่าหรือเท่ากับ 80%	1
มากกว่าหรือเท่ากับ 50%	0.5
น้อยกว่า 50%	0.1

ในตารางที่ 4 เป็นตัวอย่างของกฎตรรกศาสตร์คลุมเครือ 13 กฎจากทั้งหมด 395 กฎ ซึ่งสังเกตเห็นว่าแต่ละกฎมีตัวแปรนำเข้าเพียงตัวแปรเดียว เนื่องจากการวิจัยนี้ได้สร้างกฎโดยแยกพิจารณาการสอดคล้องกับระดับคะแนนในภาพรวมของทั้งโรงแรมเป็นเรื่องๆ ทำให้แต่ละเรื่องมีค่าน้ำหนักของกฎเฉพาะของตนเอง และอาจไม่เท่ากับเรื่องอื่นๆ ดังนั้น จึงได้สร้างกฎแต่ละกฎใน

ลักษณะที่มีตัวแปรนำเข้าเพียงตัวแปรเดียว เพื่อให้กฎที่ได้มีความเหมาะสมต่อการใช้งานด้านการวิเคราะห์บทวิจารณ์มากที่สุด

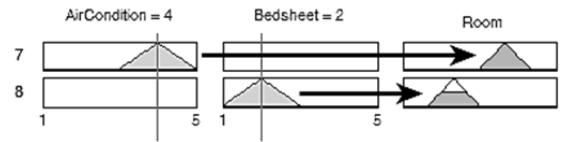
ตารางที่ 4 ตัวอย่างกฎตรรกศาสตร์คลุมเครือที่ใช้ในการคำนวณระดับคะแนนในภาพรวม

Rules	If	Then	Weight
กฎที่ใช้คำนวณคะแนนในภาพรวมของโรงแรม			
1	Room = Average	Hotel = Average	1
2	Room = Good	Hotel = Good	1
3	Service = Poor	Hotel = Poor	1
4	Bathroom = Average	Hotel = Average	1
5	Value = Excellent	Hotel = Excellent	0.5
6	Location = Excellent	Hotel = Excellent	0.1
กฎที่ใช้คำนวณคะแนนในรายละเอียดของโรงแรม			
7	AirCondition = Good	Room = Good	1
8	Bedsheet = Bad	Room = Bad	0.5
9	TVSet = Average	Room = Average	0.1
10	Employee = Poor	Service = Poor	0.5
11	Reception = Excellent	Service = Excellent	0.1
12	Bathtub = Terrible	Bathroom = Terrible	0.5
13	Toilet = Good	Bathroom = Good	0.1

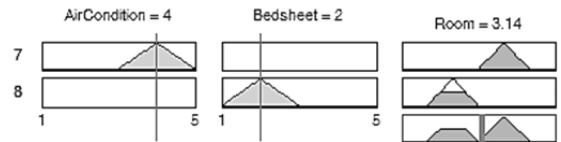
• ส่วนกลไกการอนุมาน (Inference Engine)

เป็นส่วนที่ใช้วินิจฉัยเพื่อตัดสินใจระดับคะแนนในภาพรวมตามกฎที่มีอยู่แบบลำดับขั้น ซึ่งจากกฎและน้ำหนักที่ได้ในฐานกฎพีชชีจะนำมาใช้เป็นเกณฑ์การคำนวณระดับคะแนนในภาพรวม โดยพิจารณาระดับคะแนนของทุกเรื่องในขั้นก่อนหน้าเป็นข้อมูลนำเข้า และระดับคะแนนในภาพรวมเป็นข้อมูลผลลัพธ์ โดยในลำดับขั้นแรกกลไกการอนุมานจะพิจารณาระดับคะแนนในภาพรวมจากกฎของพีชชีชุดในกลุ่มเรื่องห้องน้ำ ห้องพัก และการบริการก่อน จากนั้นจึงนำระดับคะแนนในภาพรวมที่ได้จากลำดับขั้นแรกนี้มาหาค่าคะแนนในภาพรวมของทั้งโรงแรมต่อไป

ในกลไกการอนุมาน หากคะแนนที่เป็นข้อมูลนำเข้ามีการเปลี่ยนแปลง คะแนนของข้อมูลผลลัพธ์ก็จะเปลี่ยนตามไปด้วย ทั้งนี้ การเปลี่ยนแปลงจะเกิดขึ้นมากหรือน้อยขึ้นอยู่กับน้ำหนักของกฎที่เกี่ยวข้องกับข้อมูลผลลัพธ์นั้น



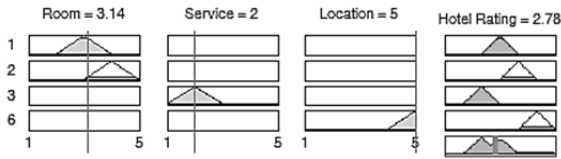
รูปที่ 6 ตัวอย่างการอนุมานค่าผลลัพธ์ เมื่อค่าน้ำหนักของกฎเป็น 1 และ 0.5



รูปที่ 7 ตัวอย่างการคำนวณระดับคะแนนในภาพรวมของเรื่องห้องพัก เมื่อคะแนนเรื่องเครื่องปรับอากาศเป็น 4 คะแนน และเรื่องผ้าปูเตียงเป็น 2 คะแนน

ตัวอย่างเช่น การคำนวณหาระดับคะแนนในภาพรวมของห้องพักที่พิจารณาจากข้อมูลนำเข้า คือคะแนนของเครื่องปรับอากาศ (AirCondition) จำนวน 4 คะแนน และผ้าปูเตียง (Bedsheet) จำนวน 2 คะแนน เมื่อนำมาอนุมานจะไปเรียกใช้กฎที่ 7 และกฎที่ 8 ดังแสดงในรูปที่ 6 จะเห็นได้ว่ากฎที่ 7 “If AirCondition = Good Then Room = Good” มีค่าน้ำหนักเท่ากับ 1 ทำให้ระดับความเป็นสมาชิกของ “Room=Good” มีค่าสูงสุด ดังแสดงในภาพด้านขวาของรูปที่ 6 ในขณะที่กฎที่ 8 “If Bedsheet = Bad Then Room = Bad” มีค่าน้ำหนักเท่ากับ 0.5 ทำให้ระดับความเป็นสมาชิกของ “Room=Bad” มีค่าเพียงครึ่งหนึ่ง เป็นต้น

• **ส่วนดีฟัซซิฟิเคชัน (Defuzzification)** เป็นส่วนที่รวมผลลัพธ์ที่ได้จากการทำงานของกฎต่างๆ และแปลงให้เป็นระดับคะแนนในภาพรวม โดยในงานวิจัยนี้เลือกใช้วิธีเซนทรอยด์ (Centroid of Area Method: COA) [15] ในการคำนวณระดับคะแนนในภาพรวม ยกตัวอย่างการคำนวณระดับคะแนนในภาพรวมของเรื่องห้องพักจากการพิจารณาจากคะแนนของเครื่องปรับอากาศจำนวน 4 คะแนน และผ้าปูเตียงจำนวน 2 คะแนน จะได้ผลลัพธ์คือ ห้องพักมีระดับคะแนนเท่ากับ 3.14 คะแนน จากการรวมผลลัพธ์ของกฎที่ 7 และกฎที่ 8 ดังแสดงในรูปที่ 7

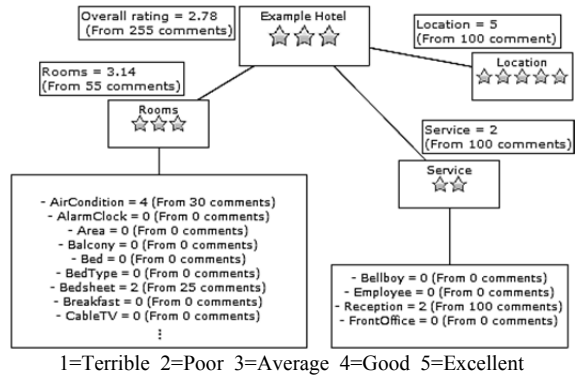


รูปที่ 8 ตัวอย่างการคำนวณระดับคะแนนในภาพรวมของ
ทั้งโรงแรม เมื่อคะแนนเรื่องห้องพักเป็น 3.14
คะแนน เรื่องการบริการเป็น 2 คะแนน และเรื่อง
ทำเลที่ตั้งเป็น 5 คะแนน

ทั้งนี้ เมื่อนำระดับคะแนนในเรื่องห้องพักจำนวน
3.14 คะแนน มาคำนวณหาระดับคะแนนในภาพรวมของ
ทั้งโรงแรม โดยพิจารณาจากคะแนนของการบริการจำนวน
2 คะแนน และเรื่องทำเลที่ตั้งจำนวน 5 คะแนนร่วมด้วย
จะได้ผลลัพธ์เป็นคะแนนในภาพรวมของทั้งโรงแรมมีค่า
2.78 คะแนน ดังแสดงในรูปที่ 8 ทั้งนี้ จะสังเกตเห็นว่า
ระดับคะแนนในเรื่องทำเลที่ตั้งสอดคล้องกับคะแนนในภาพ
รวมของทั้งโรงแรมน้อยมาก เนื่องจากค่าน้ำหนักของกฎ
ในเรื่องนี้มีค่าเพียง 0.1 ตามกฎที่ 6 ในตารางที่ 4 ซึ่ง
แตกต่างจากกฎที่ 1-3 ที่มีค่าน้ำหนักของกฎสูงสุดคือ 1

2.5 ส่วนอธิบายความรู้

เป็นการแปลงความรู้ที่อยู่ในฐานความรู้ให้อยู่ใน
รูปแบบที่ผู้ใช้สามารถเข้าใจได้ง่ายและชัดเจนมากยิ่งขึ้น
โดยความรู้ของระบบจะแสดงเป็นกราฟต้นไม้ของโรงแรม
แห่งหนึ่งตามรายละเอียดของโรงแรมที่นักท่องเที่ยวได้
แสดงความคิดเห็นไว้ และแบ่งตามหมวดหมู่ที่กำหนด
เช่น ทำเลที่ตั้ง การบริการ รายละเอียดของการบริการ
ห้องพัก และรายละเอียดของที่พัก เป็นต้น โดยการ
แบ่งหมวดหมู่ดังกล่าวได้มาจากงานวิจัยที่ได้ศึกษาความ
พึงพอใจของนักท่องเที่ยว [16]-[19] นอกจากนี้ กราฟต้นไม้
ยังแสดงระดับคะแนนและจำนวนคำวิจารณ์ของเรื่องต่างๆ
ที่สกัดได้อย่างเป็นลำดับขั้น โดยระบบจะแสดงผลระดับ
คะแนนโดยรวมเป็นรูปดาวในจำนวนที่แตกต่างกันตามเศษ
ษัตติยนิยม ซึ่งหากเศษษัตติยนิยมมีค่าอยู่ในช่วง 0.00-0.24 จะ
แสดงรูปดาวจำนวนเท่ากับจำนวนเต็มของระดับคะแนน



รูปที่ 9 ตัวอย่างการแสดงผลความรู้ที่สกัดได้

ที่คำนวณได้ หากเศษษัตติยนิยมมีค่าอยู่ในช่วง 0.25-0.74
จะแสดงรูปดาวจำนวนเท่ากับจำนวนเต็มของระดับ
คะแนนที่คำนวณได้ และเพิ่มรูปดาวอีกครึ่งดวง และ
หากเศษษัตติยนิยมมีค่าอยู่ในช่วง 0.75-0.99 จะแสดงรูปดาว
จำนวนเพิ่มขึ้น 1 ดวงจากจำนวนเต็มของระดับคะแนนที่
คำนวณได้ ตัวอย่างการแสดงผลระดับคะแนนโดยรวมเป็น
รูปดาว เมื่อจำนวนเต็มของระดับคะแนนทั้งโรงแรมเท่ากับ
4 คะแนนแสดงดังตารางที่ 5

ตารางที่ 5 ตัวอย่างการแสดงผลระดับคะแนนเป็นรูปดาว
เมื่อจำนวนเต็มของระดับคะแนนทั้งโรงแรม
เท่ากับ 4 คะแนน

เศษษัตติยนิยม	จำนวนรูปดาวที่แสดง
4.00-4.24	★★★★
4.25-4.74	★★★★½
4.75-4.99	★★★★★

จากรูปที่ 9 ได้แสดงรายละเอียดของโรงแรมแห่ง
หนึ่งว่าเป็นโรงแรม 3 ดาว (ระดับคะแนน 2.78) และม
ีการวิจารณ์ของผู้ที่เคยไปมาแล้วทั้งหมด 255 คำวิจารณ์
โดยระดับคะแนน 2.78 นั้นเกิดจากการแสดงความคิดเห็น
ในด้านต่างๆ ได้แก่ 1) ทำเลที่ตั้ง 5 คะแนน จาก 100

คำวิจารณ์ 2) การบริการ 2 คะแนน จาก 100 คำวิจารณ์ และ 3) ห้องพัก 3.14 คะแนน จาก 55 คำวิจารณ์ ซึ่งการอธิบายด้วยกราฟเพื่อแสดงรายละเอียดของโรงแรมนี้จะเป็นประโยชน์อย่างยิ่งต่อการตัดสินใจเลือกโรงแรมให้ตรงกับความต้องการของนักท่องเที่ยวได้มากที่สุด

3. การทดสอบระบบ

3.1 สภาพแวดล้อมในการทดสอบระบบ

ในการประเมินความถูกต้องของการสกัดความรู้ ข้อมูลที่นำมาทดสอบจะเป็นคนละชุดกับข้อมูลที่นำมาใช้ในการสร้างขั้นตอนวิธีในการสกัดความรู้ โดยผู้วิจัยได้สุ่มเก็บบทวิจารณ์จำนวน 100 บทวิจารณ์ จากโรงแรมจำนวน 5 โรงแรมซึ่งอยู่ในระดับความพึงพอใจตามที่ผู้ใช้ให้คะแนนต่างกัน 5 ระดับ (ระดับที่ 1-5) จำนวนระดับละ 20 บทวิจารณ์ จากเว็บไซต์ TripAdvisor.com

โดยทั่วไป งานวิจัยทางด้านการสืบค้นสารสนเทศ (Information Retrieval) จะจำแนกข้อมูลออกเป็น 4 กลุ่มเพื่อใช้ในการประเมินผล แต่ในที่นี้ได้นำมาประยุกต์ใช้กับการสกัดสารสนเทศ ซึ่งจะมีประเภทของข้อมูลที่ใช้เพื่อประเมินผลดังแสดงในตารางที่ 6 อันได้แก่ ข้อมูลที่สามารถสกัดได้และมีความเกี่ยวข้อง (True Positive: TP) ข้อมูลที่สามารถสกัดได้แต่ไม่มีความเกี่ยวข้อง (False Positive: FP) ข้อมูลที่ไม่สามารถสกัดได้และไม่มีความเกี่ยวข้อง (True Negative: TN) และข้อมูลที่ไม่สามารถสกัดได้แต่มีความเกี่ยวข้อง (False Negative: FN) [20]

ตารางที่ 6 ข้อมูล 4 ประเภทที่แตกต่างกันตามความเกี่ยวข้องและการสกัดได้

	ข้อมูลที่มีความเกี่ยวข้อง (+)	ข้อมูลที่ไม่มีความเกี่ยวข้อง (-)
ข้อมูลที่สกัดได้ (+)	TP	FP
ข้อมูลที่ไม่ได้สกัด (-)	FN	TN

ในการทดสอบความรู้ที่สกัดได้จากระบบ ผลการทดสอบจะพิจารณาถึงค่ามาตรฐานทั่วไปที่ใช้ในการประเมิน

ความถูกต้องของการสกัดสารสนเทศ รวมทั้งพิจารณาถึงระดับคะแนนในภาพรวมของทั้งโรงแรมที่ระบบสกัดได้ การทดสอบแบ่งออกเป็น 3 ประเด็น ได้แก่

ประเด็นที่ 1 การทดสอบความถูกต้องในการสกัดความรู้แบบอัตโนมัติที่ได้พัฒนาขึ้นมา โดยใช้วิธีวัดค่ามาตรฐานในการประเมินความถูกต้อง 3 ค่า ได้แก่ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) โดยมีสูตรดังสมการที่ 2, 3 และ 4 ตามลำดับ [20] ซึ่งวัตถุประสงค์ของการทดสอบในประเด็นนี้คือเพื่อประเมินความถูกต้องของขั้นตอนวิธีการสกัดความรู้ที่ได้พัฒนาขึ้น

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \times 100\% \quad (2)$$

$$Precision = \frac{TP}{(TP + FP)} \times 100\% \quad (3)$$

$$Recall = \frac{TP}{(TP + FN)} \times 100\% \quad (4)$$

ประเด็นที่ 2 การทดสอบความใกล้เคียงของระดับคะแนนในภาพรวมของทั้งโรงแรมที่ระบบสกัดได้กับที่นักท่องเที่ยวกำหนด กรณีระบบนำเทคนิคการรากลุ่มเครื่องมือในการคำนวณระดับคะแนน ซึ่งวัตถุประสงค์ของการทดสอบในประเด็นนี้คือเพื่อเปรียบเทียบความใกล้เคียงของระดับคะแนนในภาพรวมของทั้งโรงแรมที่ระบบคำนวณได้ โดยใช้ระดับคะแนนที่นักท่องเที่ยวหรือผู้กำหนดไว้เป็นเป้าหมายในการวัดความถูกต้องของระดับคะแนน

ประเด็นที่ 3 การเปรียบเทียบระดับคะแนนที่ระบบสกัดได้ กรณีใช้ตรรกศาสตร์กลุ่มเครื่องมือและไม่ใช้ตรรกศาสตร์กลุ่มเครื่องมือในการคำนวณระดับคะแนน ซึ่งวัตถุประสงค์ของการทดสอบในประเด็นนี้คือเพื่อเปรียบเทียบความใกล้เคียงของระดับคะแนนในภาพรวมของทั้งโรงแรมที่ระบบคำนวณได้ ระหว่างกรณีใช้และไม่ใช้เทคนิคการรากลุ่มเครื่องมือ โดยนำระดับคะแนนที่นักท่องเที่ยวหรือผู้กำหนดไว้มาเป็นเป้าหมายในการวัดความถูกต้องของระดับคะแนน

สำหรับกระบวนการคำนวณระดับคะแนนในภาพรวมของระบบโดยใช้เทคนิคตรรกศาสตร์คลุมเครือจะดำเนินการดังที่ได้กล่าวมาแล้วในหัวข้อที่ 2.4.2 แต่สำหรับกระบวนการคำนวณระดับคะแนนในภาพรวมของระบบโดยไม่ใช้ตรรกศาสตร์คลุมเครือจะใช้วิธีการหาค่าเฉลี่ยของระดับคะแนนเฉพาะเรื่องทุกเรื่องที่นักท่องเที่ยวยุติการณ ดึงสมการที่ 5

$$Overall\ Score = \frac{\sum_{i=1}^n Score_i}{n} \quad (5)$$

โดยที่ Overall Score หมายถึงระดับคะแนนในภาพรวมของทั้งโรงแรม Score_i หมายถึงระดับคะแนนในเรื่องวิจารณ์ที่ i และ n หมายถึงจำนวนเรื่องทั้งหมดที่วิจารณ์

3.2 ผลการทดสอบและการอภิปรายผล

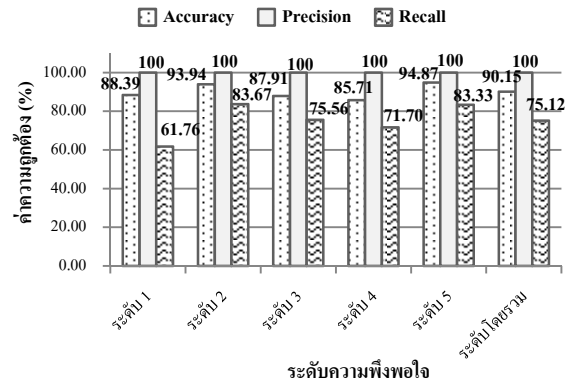
จากการจำลองสภาพแวดล้อมเพื่อทดสอบระบบสามารถอภิปรายผลโดยแยกเป็นประเด็นดังที่กล่าวมาแล้ว ดังนี้

ประเด็นที่ 1 การทดสอบความถูกต้องโดยวิธีวัดค่ามาตรฐานในการประเมิน สามารถจำแนกข้อมูลผลการทดสอบออกเป็น 4 กลุ่ม ดังปรากฏในตารางที่ 7

ตารางที่ 7 จำนวนข้อมูลที่ใช้ในการประเมินผล

ระดับ	ข้อมูลที่ให้ประเมินผล			
	TP	FP	FN	TN
1	21	0	13	78
2	41	0	8	83
3	34	0	11	46
4	38	0	15	52
5	20	0	4	54
โดยรวม	154	0	51	313

จากจำนวนข้อมูลข้างต้นสามารถนำมาคำนวณค่ามาตรฐานในการวัดความถูกต้องได้ดังรูปที่ 10 ซึ่งอธิบายได้ว่าระบบที่นำเสนอมีค่าความถูกต้องของการสกัดความรู้



รูปที่ 10 แผนภูมิสรุปผลค่ามาตรฐานในการประเมินความถูกต้องของการสกัดความรู้

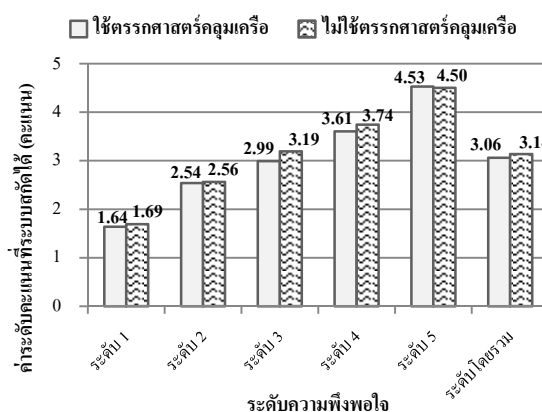
เท่ากับ 90.15% ค่าความแม่นยำเท่ากับ 100% และค่าความระลึกเท่ากับ 75.12% โดยสาเหตุที่ทำให้ค่าความแม่นยำ (Precision) ของทุกระดับความพึงพอใจมีค่าเท่ากับ 100% เสมอ เนื่องจากการศึกษาครั้งนี้ได้สกัดความรู้โดยใช้ออนโทโลยีที่มีจำนวนคำศัพท์คงที่และมีความเกี่ยวข้องเฉพาะกับเรื่องโรงแรมเท่านั้น ดังนั้น ข้อมูลที่ไม่มีความเกี่ยวข้องกับเรื่องดังกล่าวจึงไม่ถูกสกัดออกมา ทำให้ค่า FP ในทุกระดับเป็น 0 ส่วนสาเหตุที่ค่าความระลึกในระดับความพึงพอใจที่ 1 มีค่าน้อยที่สุด เนื่องจากในบทวิจารณ์ระดับที่ 1 นั้น ผู้ใช้ส่วนใหญ่จะวิจารณ์ในลักษณะบรรยายหลายประโยคมากกว่าการกล่าวถึงเรื่องที่ไม่พอใจโดยตรงเพียงประโยคเดียว ทำให้กระบวนการสกัดความรู้ที่ทำที่ละประโยคดังที่นำเสนอในงานวิจัยนี้ไม่สามารถสกัดได้อย่างถูกต้อง เมื่อเปรียบเทียบกับบทวิจารณ์ในระดับความพึงพอใจอื่นๆ ที่ผู้ใช้มักจะวิจารณ์อย่างตรงไปตรงมา จึงทำให้การสกัดที่ละประโยคมีความถูกต้องมากกว่า [21]

ประเด็นที่ 2 การทดสอบความใกล้เคียงของระดับคะแนนที่ระบบสกัดได้กับระดับคะแนนที่นักท่องเที่ยวกำหนด กรณี นำเทคนิคตรรกศาสตร์คลุมเครือมาใช้ในการคำนวณระดับคะแนน สามารถอธิบายได้ว่า ระดับคะแนนเฉลี่ยที่ระบบสกัดได้กับระดับคะแนนที่ผู้ให้กำหนดมีความแตกต่างกัน 0.06 คะแนน

(3.06-3.00) ในขณะที่เมื่อพิจารณาแยกตามกลุ่มระดับคำวิจารณ์พบว่า ระดับความพึงพอใจที่ 3 มีค่าแตกต่างกันน้อยที่สุด 0.01 คะแนน (2.99-3.00) และระดับความพึงพอใจที่ 1 มีค่าแตกต่างมากที่สุดคือ 0.64 คะแนน (1.64-1.00) ดังแสดงในรูปที่ 11

สำหรับสาเหตุที่ทำให้ระดับความพึงพอใจที่ 1 มีค่าคะแนนแตกต่างจากที่ผู้ใช้กำหนดมากที่สุด นอกจากจะเนื่องมาจากเหตุผลประการเดียวกันที่กล่าวมาแล้วในประเด็นที่ 1 ยังเกิดจากการที่ระดับคะแนนของคำศัพท์ในฐานความรู้ที่ระบบนำมาใช้คำนวณเป็นคะแนนเฉพาะเรื่องนั้น เป็นระดับคะแนนที่ได้จากการยืนยันของผู้เชี่ยวชาญด้านภาษา ซึ่งในความเป็นจริงแล้ว ผู้ใช้งานแต่ละคนอาจใช้คำศัพท์ในการแสดงความคิดเห็นที่ไม่สอดคล้องกับระดับคะแนนที่ได้จากการยืนยันของผู้เชี่ยวชาญ จึงทำให้ระดับคะแนนที่ระบบสกัดได้มีความแตกต่างจากที่ผู้ใช้กำหนด ดังจะเห็นได้ชัดเจนในระดับความพึงพอใจที่ 1 นี้ ซึ่งพบว่าระดับคะแนนเฉพาะเรื่องส่วนใหญ่ที่ระบบคำนวณได้มีค่าเท่ากับ 2 คะแนน อันมีความหมายใกล้เคียงกับระดับคะแนนเท่ากับ 1 ดังนั้น เมื่อนำคะแนนทั้งหมดมาคำนวณรวมกัน จึงทำให้ค่าระดับคะแนนที่ได้มีแนวโน้มเข้าใกล้ 2 คะแนนมากกว่า 1 คะแนน ด้วยเหตุนี้ ระดับคะแนนที่ระบบสกัดได้จึงมีความแตกต่างจากที่ผู้ใช้กำหนดมาก

ประเด็นที่ 3 การเปรียบเทียบระดับคะแนนที่ระบบสกัดได้ กรณี ใช้ตรรกศาสตร์คลุมเครือและไม่ใช่ตรรกศาสตร์คลุมเครือในการคำนวณระดับคะแนน จากรูปที่ 11 สามารถอธิบายได้ว่า ระดับคะแนนเฉลี่ยที่ระบบสกัดได้ในกรณีไม่ใช่ตรรกศาสตร์คลุมเครือในการคำนวณระดับคะแนน กับระดับคะแนนที่ผู้ใช้กำหนดมีความแตกต่างกัน 0.14 คะแนน (3.14-3.00) ซึ่งเมื่อเปรียบเทียบกับกรณีนำเทคนิคตรรกศาสตร์คลุมเครือมาใช้จะเห็นได้ว่าระดับคะแนนกรณีไม่ใช่ตรรกศาสตร์คลุมเครือมีความแตกต่างกับความพึงพอใจของนักทอ้งเที่ยวมากกว่า นอกจากนี้ เมื่อพิจารณาแยกตามกลุ่มระดับคำวิจารณ์พบว่า ระดับความพึงพอใจที่ 3 มีค่าแตกต่างกันน้อยที่สุดคือ 0.19 คะแนน (3.19-3.00) และ



รูปที่ 11 แผนภูมิสรุปผลค่าระดับคะแนนที่ระบบสกัดได้กับระดับคะแนนที่ผู้กำหนด

ระดับความพึงพอใจที่ 1 มีค่าแตกต่างมากที่สุดคือ 0.69 คะแนน (1.69-1.00) เช่นเดียวกับกรณีนำเทคนิคตรรกศาสตร์คลุมเครือมาใช้ ทั้งนี้ เมื่อนำทั้งสองกรณีมาเปรียบเทียบกันโดยพิจารณาแยกตามกลุ่มระดับคำวิจารณ์ สามารถอธิบายได้ว่ากรณีนำเทคนิคตรรกศาสตร์คลุมเครือมาใช้ทำให้ได้ระดับคะแนนที่มีความใกล้เคียงกับความพึงพอใจของผู้ใช้มากกว่ากรณีไม่ใช่เทคนิคตรรกศาสตร์คลุมเครือ ดังจะเห็นได้จากระดับความพึงพอใจที่ 1-3 และ 5 (0.64-0.69, 0.54-0.56, 0.01-0.19 และ 0.47-0.5 ตามลำดับ)

สำหรับสาเหตุที่ทำให้ระดับ 4 ในกรณีนำตรรกศาสตร์คลุมเครือมาใช้มีค่าคะแนน 3.61 แตกต่างจากความพึงพอใจของนักทอ้งเที่ยว 0.39 คะแนน ซึ่งมากกว่ากรณีไม่นำเทคนิคตรรกศาสตร์คลุมเครือมาใช้ที่มีค่าคะแนน 3.74 และแตกต่างเพียง 0.26 คะแนน เนื่องจากคำวิจารณ์ส่วนใหญ่ในระดับนี้เป็นคำวิจารณ์ในเรื่องทำเลที่ตั้ง ซึ่งผู้วิจัยได้กำหนดค่าน้ำหนักของกฎตรรกศาสตร์คลุมเครือที่ใช้ในการคำนวณระดับคะแนนไว้เป็นจำนวนน้อย ดังแสดงในกฎที่ 6 ของตารางที่ 4 จึงทำให้การวิจารณ์ถึงเรื่องทำเลที่ตั้งสอดคล้องกับระดับคะแนนในภาพรวมน้อยมาก แต่ในขณะที่การวิจารณ์ที่พบในระดับนี้กล่าวถึงทำเลที่ตั้งเป็นจำนวนมาก ด้วยเหตุนี้ เมื่อนำคะแนนในภาพรวมจึงทำให้ระดับความพึงพอใจที่ 4 กรณี

นำเทคนิคตรรกศาสตร์คลุมเครือมาใช้มีค่าคะแนนแตกต่างจากความพึงพอใจของนักท่องเที่ยวมากกว่ากรณีไม่นำเทคนิคตรรกศาสตร์คลุมเครือมาใช้

จากการทดสอบเปรียบเทียบค่าระดับคะแนนที่ระบบสกัดได้ในประเด็นที่ 2 และ 3 ดังกล่าวข้างต้น สามารถสรุปได้ว่า การนำเทคนิคตรรกศาสตร์คลุมเครือมาใช้ในการคำนวณ ทำให้ได้ผลลัพธ์ที่ใกล้เคียงกับนักท่องเที่ยวหรือผู้กำหนดมากกว่าวิธีการแบบดั้งเดิมที่ไม่นำเทคนิคตรรกศาสตร์คลุมเครือมาใช้ ซึ่งถือว่าเป็นวิธีการที่เหมาะสมและสอดคล้องกับวัตถุประสงค์ของงานวิจัยทางด้านการประมวลผลภาษาธรรมชาติ ทั้งนี้ เนื่องจากการคำนวณโดยใช้เทคนิคตรรกศาสตร์คลุมเครือได้มีการกำหนดค่าน้ำหนักที่เหมาะสมให้กับแต่ละกฎของคุณสมบัติของโรงแรม ทำให้ระดับคะแนนในภาพรวมที่คำนวณได้มีความสอดคล้องใกล้เคียงกับที่นักท่องเที่ยวหรือผู้กำหนดมากที่สุด

4. สรุป

งานวิจัยนี้ได้พัฒนาระบบสกัดความรู้แบบอัตโนมัติจากบทวิจารณ์ออนไลน์ของนักท่องเที่ยว โดยนำตรรกศาสตร์คลุมเครือมาใช้ในการคำนวณระดับคะแนนจากบทวิจารณ์ของผู้ใช้ เพื่อให้ผลลัพธ์มีความใกล้เคียงกับความพึงพอใจของนักท่องเที่ยวมากที่สุด ในการพัฒนาระบบนั้น ผู้วิจัยได้สร้างเว็บไซต์เพื่อจัดเก็บบทวิจารณ์ในฐานข้อมูล พร้อมทั้งประมวลผลที่ได้จากการสกัดความรู้ให้อยู่ในรูปแบบที่ผู้ใช้สามารถนำความรู้เหล่านั้นไปใช้ประกอบการตัดสินใจได้ทันที ผลการทดสอบพบว่า ระบบที่นำเสนอมีความถูกต้องของการสกัดความรู้เท่ากับ 90.15% ค่าความแม่นยำเท่ากับ 100% ค่าความระลึกเท่ากับ 75.12% และค่าระดับคะแนนของคำวิจารณ์โดยรวมแตกต่างจากที่นักท่องเที่ยวกำหนดเท่ากับ 0.06 คะแนน

สำหรับแนวทางในการพัฒนางานวิจัยต่อไปในอนาคต จะเห็นได้ว่าการสกัดความรู้แบบอัตโนมัตินั้น จำเป็นต้องบำรุงรักษาฐานความรู้ให้มีความสมบูรณ์และเป็นปัจจุบันอยู่เสมอ เพื่อนำมาประกอบการพัฒนาระบบให้ผลลัพธ์มีความถูกต้องมากที่สุด เช่น การเพิ่มเติม

คำศัพท์ให้มากขึ้น ทั้งในส่วน of คำที่มีความหมายเหมือนกันแต่เขียนต่างกัน คำคุณศัพท์ คำกริยาวิเศษณ์ รวมทั้งอนุประโยค และไวยากรณ์ที่ใช้ในทฤษฎีไวยากรณ์ไม่พึงบริบท เนื่องจากข้อมูลเหล่านี้ล้วนเป็นแหล่งทรัพยากรความรู้ที่ส่งผลต่อการคำนวณระดับคะแนนและการสกัดความรู้ทั้งสิ้น

ระบบสกัดความรู้แบบอัตโนมัติที่นำเสนอนี้จะมีประโยชน์อย่างยิ่ง หากผู้ประกอบการธุรกิจอิเล็กทรอนิกส์นำไปประยุกต์ใช้เก็บและวิเคราะห์ความคิดเห็นของนักท่องเที่ยวเกี่ยวกับสินค้าและบริการ แล้วนำข้อมูลเหล่านั้นมาพัฒนาสินค้าหรือบริการของตนเองให้ตอบสนองความต้องการของนักท่องเที่ยวให้มากที่สุด นอกจากนี้ระบบจะช่วยให้นักท่องเที่ยวสามารถนำความรู้ในเรื่องที่สกัดไปใช้ประกอบการตัดสินใจได้อย่างสะดวกและรวดเร็วอีกด้วย

เอกสารอ้างอิง

- [1] M. G. Miniwatts. (2010, Dec. 7). *World Internet Usage and Population Statistics*. [Online]. Available : <http://www.internetworldstats.com/stats3.htm>
- [2] World Tourism Organization, *Yearbook of Tourism Statistics: Data 2001-2005 the 59th Edition*, Madrid: WTO, 2007, pp. 8-9.
- [3] P.D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," in *Proc. the 40th Annual Meeting of the Association for Computational Linguistics*, 2002, pp. 417-424.
- [4] M. Taboada and J. Grieve, "Analyzing Appraisal Automatically," in *Proc. the AAAI Spring Symposium*, 2004, pp. 158-161.
- [5] B. Whitman and D. Ellis, "Automatic Record Reviews," in *Proc. the 5th International Society for Music Information Retrieval*, 2004, pp. 470-477.



- [6] K. Zhang, R. Narayanan, and A. Choudhary, "Voice of the Customers: Mining Online Customer Reviews for Product Feature-Based Ranking," in *Proc. the 3rd conference on Online social networks (WOSN'10)*, 2010, pp. 11.
- [7] V. Ramkumar, S. Rajasekar, and S. Swamynathan, "Scoring Products from Reviews through Application of Fuzzy Techniques," *Expert Systems with Applications*, vol. 37(10), pp. 6862-6867, Oct. 2010.
- [8] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," in *Proc. the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004, pp. 168-177.
- [9] T.L. Saaty, *Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process*. (2nd edition), Pittsburgh, PA: RWS, 2006, pp. 4-5.
- [10] P. Kitwatthanathawon., T. Angskun, and J. Angskun, "An Automatic Knowledge Extraction System from Traveler Reviews," in *Proc. the 2nd ECTI-Conference on Application Research and Development*, 2010, pp. 349-354 (in Thai).
- [11] K. Siorpaes, K. Prantner, and D. Bachlechner. (2009, Jan.15). "Class Hierarchy for the e-Tourism Ontology Version 8." [Online]. Available : <http://e-tourism.deri.at/ont/>.
- [12] G. Salton and C. Buckley, "Term Weighting Approaches in Automatic Text Retrieval," *Information Processing and Management*, vol. 24(5), pp. 513-523, 1988.
- [13] J. Fry, Linguistics 165 Class Lecture, Topic: "Context-Free Grammars for English," San José State University, Spring 2004.
- [14] H.T. Nguyen and E.A. Walker, *A First Course in Fuzzy Logic (2nd ed.)*, Boca Raton, FL: Chapman & Hall, 2000. pp. 56.
- [15] J.-S. R. Jang, C.-T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing: A computational approach to learning and machine intelligence*, Upper Saddle River, NJ: Prentice Hall, 1997, pp. 75.
- [16] T-K. Hsu, Y-F. Tsai, and H-H Wu, "The Preference Analysis for Tourist Choice of Destination: A Case Study of Taiwan," *Tourism Management*, vol. 30(2), pp. 288-297, April 2009.
- [17] M.F. Cracolici and P. Nijkamp, "The Attractiveness and Competitiveness of Tourist Destinations: A Study of Southern Italian Regions," *Tourism Management*, vol. 30(3), pp. 336-344, June 2009.
- [18] C. G-Q. Chi and H. Qu, "Examining the Structural Relationships of Destination Image, Tourist Satisfaction and Destination Loyalty: An Integrated Approach," *Tourism Management*, vol. 29(4), pp. 624-636, August 2008.
- [19] I. R. Del Bosque and H. San Martin, "Tourist Satisfaction: A Cognitive-Affective Model," *Annals of Tourism Research*, vol. 35(2), pp. 551-573, April 2008.
- [20] D. Miao, Q. Duan, H. Zhang, and N. Jiao, "Rough Set Based Hybrid Algorithm for Text Classification," *Expert Systems with Applications*, vol. 36(5), pp. 9168-9174, July 2009.
- [21] P. Kitwatthanathawon, "The Development of an Automatic Knowledge Extraction System from Online Customer Reviews," Master Thesis, School of Information Technology, Suranaree University of Technology, 2009 (in Thai).