

ระบบสกัดความรู้จากบทวิจารณ์ออนไลน์ของนักท่องเที่ยว

A Knowledge Extraction System from Online Traveler Reviews

พิชญสินี กิจวัฒนาถาวร ธรา อังสกุล และจิตติมนต์ อังสกุล*

Phichayasini Kitwatthanathawon, Thara Angskun and Jitimon Angskun*

สาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี

Abstract

Currently, most travelers use the internet to retrieve information for supporting their decision in selecting the tourist places that conform to their preferences. The most common method is the decision based on reviews of experienced tourists. However, tourists must read enormous reviews in order to select their preferred tourist places. In this research, we studied and analyzed the process of an automatic knowledge extraction for applying in a design and development of an automatic knowledge extraction system from online traveler reviews. The proposed system focused on semantic analysis from hotel reviews using the existing techniques, such as ontology as the knowledge base for knowledge extraction and storing and parser to solve the semantic analysis problem. In addition, we proposed the calculation approach for a user satisfaction level in order to provide the maximize accuracy of results. The experimental results indicated that the proposed system achieved with 90.15% of accuracy, 100% of precision, 75.12% of recall and 0.29 review points that are different from the points specified by travelers.

Keyword: Knowledge Extraction; Traveler Reviews; Ontology

บทคัดย่อ

ปัจจุบันนักท่องเที่ยวส่วนใหญ่จะใช้อินเทอร์เน็ตในการค้นคืนข้อมูลเพื่อนำมาใช้ประกอบการตัดสินใจเลือกแหล่งท่องเที่ยวให้ได้ตรงกับความต้องการ ซึ่งวิธีที่นิยมใช้กันโดยทั่วไปคือ การตัดสินใจโดยใช้ข้อมูลจากบทวิจารณ์ของนักท่องเที่ยวที่เคยไปมาแล้ว แต่อย่างไรก็ตาม การเลือกแหล่งท่องเที่ยวให้ได้ตรงกับความต้องการนั้น นักท่องเที่ยวจะต้องอ่านบทวิจารณ์ทั้งหมดซึ่งมีเป็นจำนวนมาก ในการวิจัยนี้จึงได้ศึกษาและวิเคราะห์ขั้นตอนวิธีที่เกี่ยวข้องกับการสกัดความรู้แบบอัตโนมัติ เพื่อนำมาใช้ในการออกแบบและพัฒนาระบบสกัดความรู้แบบอัตโนมัติจากบทวิจารณ์ออนไลน์ของนักท่องเที่ยว โดยมุ่งเน้นที่การวิเคราะห์ความหมายของบทวิจารณ์ของนักท่องเที่ยวเกี่ยวกับที่พัก โดยนำเอาเทคนิคต่างๆ ที่มีอยู่ในปัจจุบัน เช่น ออนโทโลยี มาใช้เป็นฐานความรู้ในการสกัดและจัดเก็บความรู้ และตัวแปลภาษา มาช่วยในการวิเคราะห์เชิงความหมาย รวมทั้งพัฒนาวิธีการคำนวณระดับความพึงพอใจขึ้นมาใช้ เพื่อให้ได้มาซึ่งผลลัพธ์ที่มีความถูกต้องมากที่สุด ผลการทดลองพบว่า ระบบที่นำเสนอมีค่าความถูกต้องของการสกัดความรู้เท่ากับ 90.15% ค่าความแม่นยำเท่ากับ 100% ค่าความระลึกเท่ากับ 75.12% และค่าระดับความพึงพอใจของคำวิจารณ์โดยรวมแตกต่างจากระดับความพึงพอใจที่นักท่องเที่ยวกำหนดเท่ากับ 0.29 คะแนน

* ผู้เขียนที่ให้การติดต่อ โทร. +66 4422 4309; Fax: +66 4422 4205
E-mail address: jitimon@.sut.ac.th

บทนำ

ในปัจจุบัน การพัฒนาทางด้านเทคโนโลยีสารสนเทศและการสื่อสารมีความเจริญก้าวหน้าเป็นอย่างมาก ส่งผลให้อินเทอร์เน็ตได้เข้ามามีบทบาทในชีวิตประจำวันของผู้คนมากยิ่งขึ้น โดยทั่วโลกมีผู้ใช้งานอินเทอร์เน็ตสูงถึงประมาณ 1,966 ล้านคน (คิดเป็นร้อยละ 28.7 ของประชากรทั้งโลก) (Miniwatts, 2010) ในขณะที่ประเทศไทยก็มีผู้ใช้งานอินเทอร์เน็ตถึงประมาณ 17.5 ล้านคน (Miniwatts, 2010) จากจำนวนการใช้งานอินเทอร์เน็ตที่สูงมากดังกล่าว ได้ส่งผลให้การดำเนินกิจกรรมทางธุรกิจแบบดั้งเดิมเปลี่ยนรูปแบบไปเป็นธุรกิจอิเล็กทรอนิกส์ (E-business) มากขึ้น

ธุรกิจการท่องเที่ยวนับเป็นธุรกิจหลักอันหนึ่งของประเทศไทย ซึ่งปัจจุบันการดำเนินการขององค์กรต่างๆ ที่มีส่วนเกี่ยวข้องกับธุรกิจนี้ได้มีการปรับเปลี่ยนองค์กรของตนเองไปสู่ธุรกิจอิเล็กทรอนิกส์มากยิ่งขึ้น เพื่อเป็นแรงจูงใจให้นักท่องเที่ยวมาซื้อสินค้าหรือบริการของตน และอำนวยความสะดวกแก่นักท่องเที่ยวในการบริการจองโรงแรมที่พัก จองรถ หรือจองตั๋วเครื่องบิน ตลอดจนช่วยให้นักท่องเที่ยวสามารถสืบค้นข้อมูลด้านการท่องเที่ยว เพื่อนำมาใช้ประกอบการตัดสินใจเลือกแหล่งท่องเที่ยวให้ได้ตรงกับความต้องการอีกด้วย โดยวิธีที่นิยมใช้กันโดยทั่วไป คือ การตัดสินใจโดยใช้ข้อมูลจากบทวิจารณ์ของนักท่องเที่ยวที่เคยไปมาแล้ว ซึ่งแนวโน้มดังกล่าวได้สอดคล้องกับผลการศึกษาขององค์การส่งเสริมการท่องเที่ยวโลก (World Tourism Organization: WTO) ที่พบว่าปัจจัยสำคัญที่ส่งผลให้การท่องเที่ยวโลกเพิ่มสูงขึ้น คือ ระบบข้อมูลข่าวสารทันสมัยและก้าวหน้ามากขึ้น เพราะเมื่อนักท่องเที่ยวมีโอกาสรับทราบข้อมูลข่าวสารด้านการท่องเที่ยวมากขึ้น จะทำให้นักท่องเที่ยวมีความสนใจและมั่นใจในการเดินทางเพิ่มมากขึ้นตามไปด้วย (World Tourism Organization, 2007)

อย่างไรก็ตาม ถึงแม้ว่าการจัดเก็บข้อมูลต่างๆ จากนักท่องเที่ยวจะมีการปฏิบัติกันอยู่แล้วอย่างกว้างขวางในโลกธุรกิจอิเล็กทรอนิกส์ แต่การจะนำข้อมูลเหล่านั้นมาใช้ประโยชน์กลับเป็นเรื่องที่ยุ่ยากและต้องใช้เวลานาน เนื่องจากข้อมูลที่จัดเก็บไว้นั้นเป็นข้อมูลดิบที่ยังไม่ได้ผ่านการคัดกรอง การประมวลผล หรือการสกัด เพื่อให้ได้ความรู้ที่สามารถนำไปใช้ประกอบการตัดสินใจได้ทันที ตัวอย่างเช่น เว็บไซต์สำหรับการท่องเที่ยวที่มีการจัดเก็บบทวิจารณ์เกี่ยวกับแหล่งท่องเที่ยว ที่พัก หรือร้านอาหารในแหล่งท่องเที่ยวที่เคยไปมาแล้ว เพื่อเป็นข้อมูลให้กับนักท่องเที่ยวคนอื่นๆ ซึ่งยังไม่มีความต้องการที่ชัดเจนว่าจะไปท่องเที่ยวที่ใด และต้องการค้นหาข้อมูลเพื่อนำไปใช้ในการประกอบการวางแผนท่องเที่ยวของตน เว็บไซต์ประเภทนี้จะเพียงให้ข้อมูลอันดับความนิยมของแหล่งท่องเที่ยวโดยรวมที่นักท่องเที่ยวคนอื่นได้เคยแสดงความคิดเห็นด้วยวิธีการลงคะแนนเท่านั้น หากนักท่องเที่ยวคนใดต้องการรายละเอียดเพิ่มเติม จะต้องตามอ่านในทุกๆ ความคิดเห็นหรือทุกๆ บทวิจารณ์เอง แล้วจึงวิเคราะห์หรือตัดสินใจว่าแหล่งท่องเที่ยวนั้นควรไปหรือไม่ อีกทั้งเนื่องจากเว็บไซต์เป็นแบบตายตัว กล่าวคือ เป็นเว็บไซต์ที่ไม่สามารถโต้ตอบกับผู้ใช้งานได้ ทำให้ผู้ใช้ไม่สามารถเลือกดูข้อมูลได้ตามความต้องการของตนเอง อีกทั้งข้อมูลที่จัดเก็บไว้อย่างเป็นข้อมูลดิบที่ไม่ได้ผ่านการประมวลผล ทำให้เป็นข้อมูลที่ยังไม่ละเอียดเพียงพอสำหรับความต้องการนักท่องเที่ยว

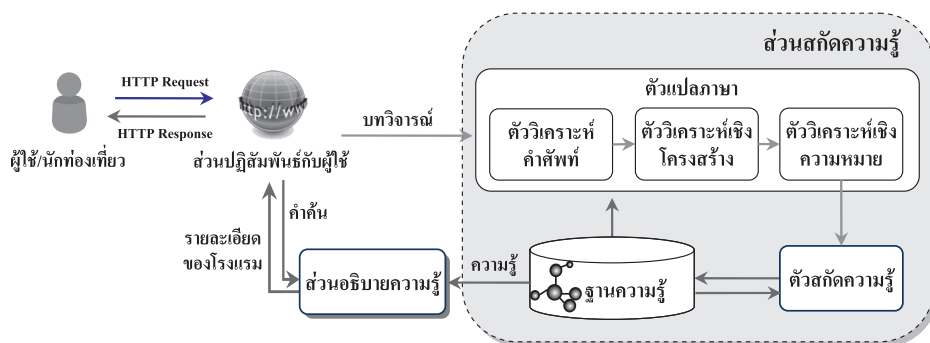
นอกจากนี้ การทบทวนวรรณกรรมที่ศึกษาเกี่ยวกับบทวิจารณ์ของลูกค้าพบว่า งานวิจัยส่วนใหญ่เป็นการจำแนกบทวิจารณ์ให้อยู่ในหมวดหมู่ที่ต้องการแบบอัตโนมัติ ตัวอย่างเช่น การสร้างขั้นตอนวิธีอย่างง่ายในการแยกประเภทคำวิจารณ์ว่ามีความหมายเชิงแนะนำหรือไม่ โดยการพยากรณ์จากการคำนวณค่าเฉลี่ยของกลุ่มคำคุณศัพท์หรือคำกริยาวิเศษณ์ (Turney, 2002) การจำแนกคุณสมบัติที่โดดเด่นของผลิตภัณฑ์และพัฒนาวิธีการจำแนกบทวิจารณ์ในแง่บวกและแง่ลบแบบอัตโนมัติ โดยในส่วนของ การสกัดรูปลักษณะของผลิตภัณฑ์และสกัดระดับความพึงพอใจ ได้นำเทคนิคด้านการสืบค้นสารสนเทศมาใช้ (Dave, Lawrence and Pennock, 2003) การจำแนกข้อความแบบอัตโนมัติโดยนำวิธีมาตรฐานในการคำนวณความหมายจากตำแหน่งของคำคุณศัพท์ในข้อความมาประยุกต์ใช้ (Taboada and Grieve, 2004) การวิเคราะห์บทวิจารณ์ทางด้านดนตรี และพัฒนาระบบที่มีกลไกในการกำจัดอคติและการบรรยายที่ไม่เกี่ยวข้องกับดนตรีของผู้วิจารณ์ โดยการเรียนรู้จากบทวิจารณ์จำนวนมาก (Whitman and Ellis, 2004) และการนำขั้นตอนวิธีให้คำแนะนำเพื่อการเรียนรู้ของเครื่องคอมพิวเตอร์มาจำแนกทัศนคติของบทวิจารณ์ภาษาจีน (Zheng and Ye, 2009)

สำหรับงานวิจัยที่ศึกษาเกี่ยวกับการสกัดความรู้จากบทวิจารณ์ของลูกค้า ได้แก่ งานวิจัยของฮูและหลู (Hu and Liu, 2004) ที่ได้สกัดความรู้จากบทวิจารณ์โดยการสรุปสาระสำคัญและนำเสนอทั้งในด้านบวกและลบให้กับลูกค้า โดยใช้วิธีการทำเหมืองข้อมูลและการประมวลผลภาษาธรรมชาติ แต่อย่างไรก็ตาม ลูกค้าไม่สามารถทราบถึงจำนวนคำวิจารณ์ต่างๆ อย่างละเอียดว่ามีผู้ที่แสดงความคิดเห็นทั้งด้านบวกหรือด้านลบในแต่ละเรื่องจำนวนเท่าใด ซึ่งในการจัดลำดับความสำคัญนั้น จำนวนของการแสดงความคิดเห็นนับว่ามีผลอย่างมากต่อการจัดลำดับความสำคัญ หรือการตัดสินใจเลือกสินค้าและบริการของลูกค้า (Saaty, 2006) ดังนั้น งานวิจัยนี้จึงนำเสนอระบบสกัดความรู้จากบทวิจารณ์ออนไลน์ของนักท่องเที่ยวอย่างละเอียด โดยมีการแสดงจำนวนความคิดเห็นของแต่ละประโยคคำวิจารณ์ และนำเสนอเทคนิคต่างๆ ที่มีอยู่ในปัจจุบันเข้ามาประยุกต์ใช้ เช่น ออนโทโลยี มาใช้เป็นฐานความรู้ในการวิเคราะห์คำศัพท์ การสกัดและจัดเก็บความรู้ นอกจากนี้ยังนำตัวแปลภาษามาช่วยในการวิเคราะห์เชิงโครงสร้าง รวมทั้งพัฒนาวิธีการคำนวณระดับคะแนนความพึงพอใจขึ้นมาใช้ เพื่อให้ได้มาซึ่งผลลัพธ์ที่มีความถูกต้องมากที่สุด โดยใช้ข้อมูลด้านการท่องเที่ยว อันได้แก่ แหล่งที่พักประเภทโรงแรม เป็นกรณีศึกษา

กรอบการทำงานของระบบ

ในการพัฒนาระบบสกัดความรู้แบบอัตโนมัติจากบทวิจารณ์ออนไลน์ของนักท่องเที่ยว มุ่งเน้นการพัฒนาขั้นตอนวิธีในการประมวลความหมายอย่างถูกต้องจากบทวิจารณ์ภาษาอังกฤษของนักท่องเที่ยวที่ได้เข้าใช้บริการโรงแรมมาแล้ว และมีความถูกต้องตามหลักไวยากรณ์ภาษาอังกฤษเท่านั้น โดยการสกัดเอาส่วนที่สำคัญของบทวิจารณ์นำมาเป็นความรู้ที่เก็บอยู่ในฐานความรู้ของระบบ และเมื่อนักท่องเที่ยวคนใหม่เข้ามาค้นหาข้อมูลก็จะดึงความรู้ที่เก็บอยู่ในฐานนั้นส่งไปให้ โดยกรอบการ

ทำงานของระบบสกัดความรู้จากบทวิจารณ์ออนไลน์ของนักท่องเที่ยว นั้น แสดงดังรูปที่ 1 ซึ่งประกอบด้วย 5 ส่วน ได้แก่ ส่วนปฏิสัมพันธ์กับผู้ใช้ ฐานความรู้ ตัวแปลภาษา ตัวสกัดความรู้ และส่วนอธิบายความรู้ โดยสามารถอธิบายอย่างละเอียดได้ดังต่อไปนี้



ภาพที่ 1 กรอบการทำงานของระบบสกัดความรู้จากบทวิจารณ์ออนไลน์ของนักท่องเที่ยว

1) **ส่วนปฏิสัมพันธ์กับผู้ใช้** ส่วนปฏิสัมพันธ์กับผู้ใช้เป็นส่วนโต้ตอบกับผู้ใช้หรือนักท่องเที่ยว ซึ่งใช้ระบบผ่านทางเว็บเบราว์เซอร์ โดยผู้ใช้จะป้อนบทวิจารณ์เกี่ยวกับโรงแรมที่เคยไปมาแล้วเข้าสู่ระบบ จากนั้นระบบจะทำการสกัดความรู้จากบทวิจารณ์ที่ป้อนเข้ามานั้นเก็บไว้ในฐานความรู้ของระบบ เมื่อผู้ใช้หรือนักท่องเที่ยวคนอื่นต้องการทราบข้อมูลเกี่ยวกับโรงแรมต่างๆ ว่าโรงแรมนั้นดีหรือไม่ ก็จะใส่คำค้นเข้าไป หากมีข้อมูลที่ตรงกับคำค้นดังกล่าว ระบบจะนำรายละเอียดของโรงแรมมาแสดงให้ผู้ใช้ทราบในรูปแบบของกราฟ ซึ่งจะอธิบายอย่างละเอียดต่อไปในส่วนอธิบายความรู้ แต่หากไม่มีข้อมูลที่ตรงกับคำค้น ระบบจะแสดงข้อความว่าไม่พบข้อมูลโรงแรมที่ค้นหา

2) **ฐานความรู้** เป็นส่วนที่ใช้เก็บข้อมูลที่เป็นประโยชน์ต่อการสกัดความรู้แบบอัตโนมัติ ตลอดจนความรู้ที่สกัดได้ ซึ่งฐานความรู้ประกอบด้วย 4 ส่วนหลักๆ ดังต่อไปนี้

- **ออนโทโลยีด้านการท่องเที่ยว** ได้ปรับปรุงจากออนโทโลยีการท่องเที่ยวออนไลน์ ชุดที่ 8 ของเสียร์พีส์ ฟรานท์เนอร์ และบาชเลชเนอร์ (Siorpaes, Prantner and Bachlechner, 2004) ซึ่งพัฒนาด้วยภาษา OWL โดยนำบทวิจารณ์ของนักท่องเที่ยวจำนวน 400 บทวิจารณ์มาวิเคราะห์เพื่อหาคำสำคัญที่อยู่ในขอบเขตที่สนใจ โดยข้อมูลบทวิจารณ์ทั้งหมดที่นำมาวิเคราะห์สามารถจำแนกตามระดับความพึงพอใจโดยรวมได้เป็น 5 กลุ่ม คือ ระดับความพึงพอใจที่ 1-5 จำนวนระดับละ 80 บทวิจารณ์ จากนั้นจึงนำคำสำคัญที่วิเคราะห์ได้มาเพิ่มเติมลงในออนโทโลยีมาตรฐานที่มีอยู่

- **คลังคำเหมือน** เป็นแหล่งรวบรวมคำต่างๆ ที่มีความหมายเหมือนกัน แต่เขียนต่างกัน ตัวอย่างเช่น คำว่า “Air Conditioner” นักท่องเที่ยวแต่ละคนอาจใช้คำที่ต่างกัน เช่น “a.c.”, “a/c”, “air condition” ทั้งนี้ จะมีการกำหนดคำที่เป็นตัวแทนของกลุ่มคำเหมือนแต่ละกลุ่มเพื่อใช้ในการแปลงคำศัพท์ให้อยู่ในรูปแบบที่เหมาะสมสำหรับการสกัดความรู้

• **คลังคำศัพท์** เป็นแหล่งรวบรวมคำศัพท์ต่างๆ โดยในงานวิจัยนี้ได้จำแนกคำศัพท์ในคลังออกเป็น 3 ประเภทใหญ่ๆ ได้แก่ 1) กลุ่มคำคุณศัพท์ ซึ่งมีการกำหนดระดับคะแนนความพึงพอใจไว้ตายตัว; 2) กลุ่มคำกริยาวิเศษณ์ ซึ่งกำหนดการเพิ่ม-ลดระดับคะแนนความพึงพอใจไว้ เมื่อนำคำกลุ่มนี้ไปประมวลผลจะทำให้ระดับความพึงพอใจที่สกัดได้มีผลลัพธ์ไปในทางเดียวกัน กล่าวคือ หากผู้ใช้วิจารณ์ในแง่ดี ระดับความพึงพอใจที่สกัดได้ก็จะยิ่งเพิ่มมากขึ้น; และ 3) กลุ่มคำกริยาวิเศษณ์แบบพิเศษ ซึ่งกำหนดการเพิ่ม-ลดระดับคะแนนความพึงพอใจไว้เช่นเดียวกับกลุ่มคำกริยาวิเศษณ์ แต่แตกต่างกันตรงที่เมื่อนำคำกลุ่มนี้ไปประมวลผลจะทำให้ระดับความพึงพอใจที่สกัดได้มีผลลัพธ์ไปในทางตรงกันข้าม กล่าวคือ หากผู้ใช้วิจารณ์ในแง่ดี ระดับความพึงพอใจที่สกัดได้ก็จะถูกลดทอนลง

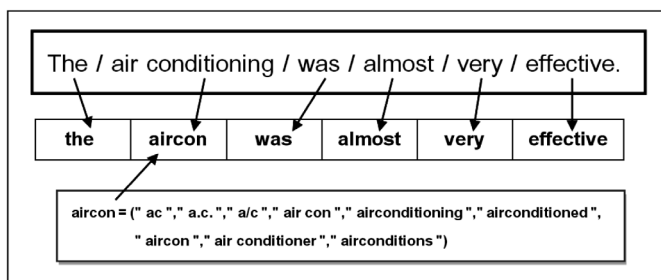
ตัวอย่างคำคุณศัพท์และระดับคะแนนความพึงพอใจที่กำหนด เช่น “Effective” = 4 และ “Bad” = 2 ตัวอย่างคำกริยาวิเศษณ์และการเพิ่ม-ลดระดับคะแนนความพึงพอใจ เช่น “Very” = +1 และ “Really” = +1 และตัวอย่างคำกริยาวิเศษณ์แบบพิเศษ เช่น “Not” = +2 และ “Almost” = +1 ทั้งนี้ การให้ระดับคะแนนความพึงพอใจของคำได้จากการยืนยันของผู้เชี่ยวชาญด้านภาษา และคะแนนแต่ละระดับจะแสดงถึงความพึงพอใจที่แตกต่างกัน โดยระดับคะแนนที่ 1 หมายถึง แย่มาก (Terrible) ระดับคะแนนที่ 2 หมายถึง แย่ (Poor) ระดับคะแนนที่ 3 หมายถึง ปานกลาง (Average) ระดับคะแนนที่ 4 หมายถึง ดี (Good) และระดับคะแนนที่ 5 หมายถึง ดีมาก (Excellent)

• **กฎไวยากรณ์ไม่พึ่งบริบท (Context Free Grammar)** เป็นกลุ่มของกฎไวยากรณ์ที่เขียนขึ้นโดยแสดงเป็นสัญลักษณ์ทางภาษา (Fry, 2004) เพื่อใช้สำหรับตรวจสอบไวยากรณ์ของประโยคคำวิจารณ์ในขั้นตอนการแปลภาษา

3) **ตัวแปลภาษา** เป็นส่วนสำคัญของระบบที่ทำให้ทราบกลุ่มคำสำคัญที่ใช้ในการวิเคราะห์ความหมายทั้งหมด สามารถแบ่งการทำงานออกเป็น 3 ส่วน ดังต่อไปนี้

• **ตัววิเคราะห์คำศัพท์ (Lexical Analyzer)** ทำหน้าที่ในการตัดคำตามช่องว่าง เพื่อแยกคำศัพท์ต่างๆ ออกจากข้อความบทวิจารณ์ของนักท่องเที่ยว ดังตัวอย่างที่แสดงในรูปที่ 2 จากนั้นจึงแปลงคำศัพท์ให้อยู่ในรูปแบบที่เหมาะสมสำหรับการสกัดความรู้ อันได้แก่ การตรวจสอบคำผิดเพื่อเปลี่ยนให้เป็นคำที่ถูกต้อง และการตรวจสอบคำที่มีความหมายเหมือนกัน แต่เขียนต่างกัน โดยเปรียบเทียบจากคลังคำเหมือนที่อยู่ในฐานความรู้ของระบบ เพื่อเปลี่ยนให้คำเหล่านั้นเป็นคำเดียวกันทั้งหมด ดังคำว่า “air conditioning” ในตัวอย่างซึ่งสามารถใช้คำได้หลายแบบ ระบบจึงเปลี่ยนคำดังกล่าวเป็นคำว่า “aircon” ซึ่งเป็นคำแทนของกลุ่มคำที่มีความหมายเหมือนกัน แต่เขียนต่างกันใน

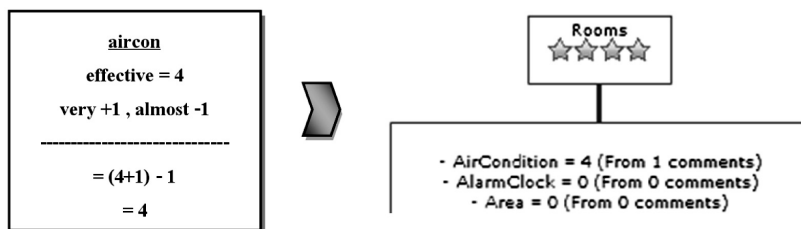
• **ตัววิเคราะห์เชิงโครงสร้าง (Syntactic Analyzer)** ทำหน้าที่แยกแยะส่วนของประโยคคำวิจารณ์ โดยการจำแนกว่าส่วนใดคือ ประธาน กริยา กรรม และส่วนขยาย ในส่วนนี้ผู้วิจัยได้นำออนโทโลยีด้านการท่องเที่ยวเข้ามาช่วยวิเคราะห์หาคำนามซึ่งเป็นคำสำคัญที่บ่งบอกถึงเรื่องที่นักท่องเที่ยวกล่าวถึง และพิจารณาเฉพาะบางประโยคของบทวิจารณ์ที่มีคำนามที่สำคัญเท่านั้น จากนั้นระบบจะจำแนกประเภทของคำศัพท์ทั้งหมดเฉพาะที่อยู่ในประโยคคำวิจารณ์นั้น และวิเคราะห์หาความเกี่ยวข้องสัมพันธ์กันระหว่างคำในประโยคตามทฤษฎีไวยากรณ์ไม่พึ่งบริบท



ภาพที่ 2 ตัวอย่างการทำงานของตัววิเคราะห์คำศัพท์

• ตัววิเคราะห์เชิงความหมาย (Semantic Analyzer) ทำหน้าที่วิเคราะห์ความหมายของคำที่ได้จากตัววิเคราะห์เชิงโครงสร้างซึ่งทราบความเกี่ยวข้องสัมพันธ์กันในประโยคคำวิจารณ์แล้ว ในการวิเคราะห์ความหมายนั้น ระบบจะให้ความสำคัญระหว่างคู่ของคำนามกับคำคุณศัพท์และคำกริยาวิเศษณ์ที่พบในประโยคคำวิจารณ์เดียวกัน โดยระบบจะค้นหาระดับคะแนนความพึงพอใจของคำคุณศัพท์และคำกริยาวิเศษณ์จากคลังคำศัพท์ จากตัวอย่างคำว่า “effective” มีค่าระดับคะแนนเท่ากับ 4 คำว่า “very” มีค่าการเพิ่ม-ลดระดับคะแนนเท่ากับ +1 และคำว่า “almost” มีค่าการเพิ่ม-ลดระดับคะแนนเท่ากับ +1 ทั้งนี้ เพื่อเตรียมข้อมูลสำหรับการคำนวณระดับความพึงพอใจทั้งในเรื่องย่อๆ และภาพรวมของบทวิจารณ์ของนักท่องเที่ยวคนนั้น

4) **ตัวสกัดความรู้** เป็นการคำนวณระดับความพึงพอใจตามคำคุณศัพท์และคำกริยาวิเศษณ์ในเรื่องที่คำนามแสดงถึง สำหรับการเพิ่มหรือลดระดับคะแนนความพึงพอใจของคำกริยาวิเศษณ์ จะพิจารณาจากค่าระดับคะแนนของคำคุณศัพท์ใน 2 กรณี ได้แก่ (1) กรณีวิจารณ์ในแง่ไม่ดี ระดับคะแนนของคำคุณศัพท์มีค่าน้อยกว่า 3 คะแนน และ (2) กรณีวิจารณ์ในแง่ดี ระดับคะแนนของคำคุณศัพท์มีค่าตั้งแต่ 3 คะแนนขึ้นไป ซึ่งทั้ง 2 กรณีนี้จะป้อนเข้าในการเพิ่มหรือลดระดับคะแนนความพึงพอใจของคำกริยาวิเศษณ์ทั้ง 2 ประเภทที่พบในประโยค ดังตัวอย่าง คำว่า “effective” ซึ่งมีค่าเท่ากับ 4 คะแนน เมื่อพิจารณาคำว่า “very” พบว่าเป็นคำกริยาวิเศษณ์ที่เมื่อนำไปประมวลผลจะทำให้ระดับความพึงพอใจมีผลไปในทางเดียวกัน กล่าวคือ กรณีนี้ผู้ใช้วิจารณ์ในแง่ดี ระดับความพึงพอใจที่สกัดได้ก็จะยิ่งเพิ่มมากขึ้น ดังนั้น คำว่า “very” จึงมีค่าเท่ากับ +1 คะแนน ต่อมาเมื่อพิจารณาคำว่า “almost” พบว่าเป็นคำกริยาวิเศษณ์แบบพิเศษที่เมื่อนำคำกลุ่มนี้ไปประมวลผลจะทำให้ระดับความพึงพอใจมีผลไปในทางตรงกันข้าม กล่าวคือ กรณีนี้ผู้ใช้วิจารณ์ในแง่ดี ระดับความพึงพอใจที่สกัดได้ก็จะถูกลดทอนลง ดังนั้น คำว่า “almost” จึงมีค่าเท่ากับ -1 คะแนน เมื่อนำระดับคะแนนความพึงพอใจของคำทั้งหมดมาคำนวณจะได้เป็นระดับความพึงพอใจเฉพาะเรื่อง ซึ่งค่าของระดับความพึงพอใจนี้จะถูกจัดเก็บลงในออนโทโลยีโดยจำแนกตามเรื่องที่วิจารณ์ ดังรูปที่ 3 ซึ่งเป็นกรณีเครื่องปรับอากาศ

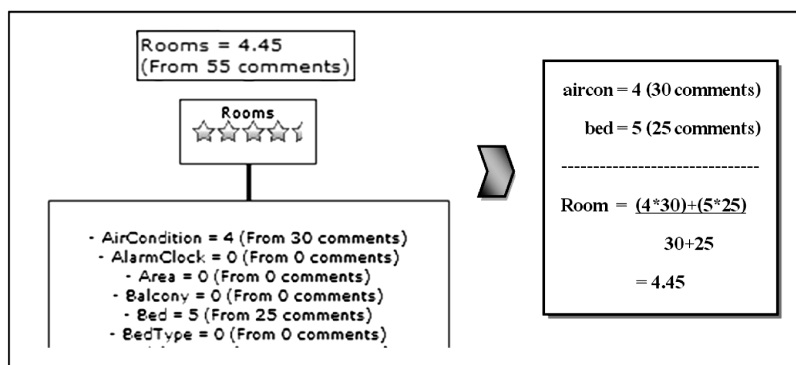


ภาพที่ 3 ตัวอย่างการคำนวณระดับความพึงพอใจเฉพาะเรื่อง

นอกจากนี้ ระบบจะประมวลรวมระดับความพึงพอใจในทุกเรื่องที่เคยสกัดได้ และแสดงเป็นระดับความพึงพอใจในภาพรวมของระดับชั้นบนถัดไป ดังรูปที่ 4 โดยจะประมวลรวมระดับความพึงพอใจไปเรื่อยๆ จนกลายเป็นระดับความพึงพอใจในภาพรวมของทั้งโรงแรม ตัวอย่างเช่น เมื่อมีผู้วิจารณ์เรื่องเครื่องปรับอากาศ จำนวน 30 คำวิจารณ์ โดยมีระดับความพึงพอใจเฉลี่ยเท่ากับ 4 คะแนน และมีผู้วิจารณ์เรื่องเตียงนอน จำนวน 25 คำวิจารณ์ โดยมีระดับความพึงพอใจเฉลี่ยเท่ากับ 5 คะแนน ซึ่งทั้ง 2 กรณีเป็นเรื่องเกี่ยวกับห้องพักทั้งสิ้น ดังนั้น ระบบจะนำข้อมูลทั้ง 2 เรื่องมาคำนวณเป็นระดับความพึงพอใจของห้องพัก (RoomScore) ตามสมการที่ 1

$$RoomScore = \frac{\sum_{i=1}^k (S_i * N_i)}{\sum_{i=1}^k N_i} \quad (1)$$

โดย S_i คือ ระดับความพึงพอใจเฉพาะเรื่อง และ N_i คือ จำนวนคำวิจารณ์เฉพาะเรื่อง ซึ่ง $1 \leq i \leq k$ และ k คือ จำนวนเรื่องทั้งหมดที่สกัดได้ ทั้งนี้ เรื่องที่สกัดได้ทั้งหมด พร้อมระดับความพึงพอใจเฉพาะเรื่อง และระดับความพึงพอใจในภาพรวม จะเป็นความรู้ที่ถูกจัดเก็บลงในฐานความรู้ของระบบ

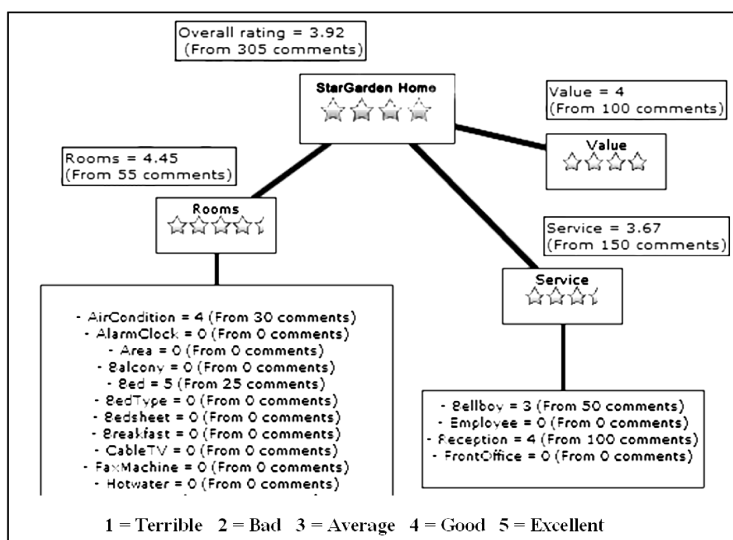


ภาพที่ 4 ตัวอย่างการคำนวณระดับความพึงพอใจในภาพรวม

ในส่วนของการแสดงผลระดับความพึงพอใจโดยรวมที่ระบบคำนวณได้นั้น จะแสดงเป็นรูปดาว ในจำนวนที่แตกต่างกันออกไปตามเลขทศนิยม โดยหากเลขทศนิยมมีค่าอยู่ในช่วง 0.00-0.24 จะแสดงรูปดาวจำนวนเท่ากับระดับความพึงพอใจที่คำนวณได้; หากเลขทศนิยมมีค่าอยู่ในช่วง 0.25-0.74 จะแสดงรูปดาวจำนวนเท่ากับระดับความพึงพอใจที่คำนวณได้ และเพิ่มรูปดาวอีกครั้งดวง; และหากเลขทศนิยมมีค่าอยู่ในช่วง 0.75-0.99 จะแสดงรูปดาวจำนวนเพิ่มขึ้น 1 ดวงจากระดับความพึงพอใจที่คำนวณได้ ทั้งนี้ รูปดาวที่แสดงไม่มีความเกี่ยวข้องกับการจัดระดับโรงแรมที่ใช้ในปัจจุบัน

5) ส่วนอธิบายความรู้ เป็นส่วนที่แปลงความรู้จากฐานความรู้ให้อยู่ในรูปแบบที่ผู้ใช้สามารถเข้าใจได้ง่ายและชัดเจนมากยิ่งขึ้น โดยความรู้ของระบบจะแสดงเป็นกราฟต้นไม้ของโรงแรมแห่งหนึ่งตามรายละเอียดของโรงแรมที่นักท่องเที่ยวได้แสดงความคิดเห็นไว้ และแบ่งตามหมวดหมู่ที่ได้กำหนด เช่น ความคุ้มค่า การบริการ รายละเอียดของการบริการ ห้องพัก และรายละเอียดของที่พัก เป็นต้น โดยการแบ่งหมวดหมู่ดังกล่าวได้มาจากการวิจัยที่ได้ศึกษาความพึงพอใจของนักท่องเที่ยว (Hsu, Tsai and Wu, 2009; Cracolici and Nijkamp, 2008; Chi and Qu, 2008; Del Bosque and San Martin, 2008) นอกจากนี้ กราฟต้นไม้ยังแสดงระดับความพึงพอใจและจำนวนคำวิจารณ์ของเรื่องต่างๆ ที่สกัดได้เป็นอย่างดีเป็นลำดับชั้น ดังแสดงในรูปที่ 5

จากรูปได้แสดงรายละเอียดของโรงแรมสตาร์การ์เด้นโฮม (StarGarden Home) ว่าเป็นโรงแรมในระดับ 4 ดาว (ระดับความพึงพอใจ 3.92) และมีการวิจารณ์ของผู้ที่เคยไปมาแล้วทั้งหมด 305 คำวิจารณ์ โดยระดับความพึงพอใจ 3.92 นั้นเกิดจากการแสดงความคิดเห็นในด้านต่างๆ ได้แก่ 1) ความคุ้มค่า (Value) 4 คะแนน จาก 100 คำวิจารณ์; 2) การบริการ (Service) 3.67 คะแนน จาก 150 คำวิจารณ์; และ 3) ห้องพัก (Rooms) 4.45 คะแนน จาก 55 คำวิจารณ์ ซึ่งการอธิบายด้วยกราฟเพื่อแสดงรายละเอียดของโรงแรมนี้จะเป็นประโยชน์อย่างยิ่งต่อการตัดสินใจในการเลือกโรงแรมให้ตรงกับความต้องการของนักท่องเที่ยวได้มากที่สุด



ภาพที่ 5 ตัวอย่างการแสดงผลความรู้ที่สกัดได้

การทดสอบระบบ

สภาพแวดล้อมในการทดสอบระบบ

ในการประเมินความถูกต้องของการสกัดความรู้ นั้น ข้อมูลที่นำมาทดสอบจะเป็นคนละชุดกับข้อมูลที่ใช้ในการสร้างขั้นตอนวิธีในการสกัดความรู้ โดยผู้วิจัยได้สุ่มเก็บบทวิจารณ์จำนวน 100 บทวิจารณ์ จากโรงแรมจำนวน 5 โรงแรมซึ่งอยู่ในระดับความพึงพอใจที่ต่างกัน 5 ระดับ จำนวนระดับละ 20 บทวิจารณ์ จากเว็บไซต์ TripAdvisor.com โดยบทวิจารณ์เหล่านี้ ผู้วิจัยได้จำแนกเป็นข้อมูลคำวิจารณ์ที่เกี่ยวข้องกับเรื่องโรงแรมจำนวน 205 คำวิจารณ์ ข้อมูลคำวิจารณ์ที่ไม่เกี่ยวข้องกับเรื่องโรงแรมจำนวน 313 คำวิจารณ์ และข้อมูลคำวิจารณ์ที่ผิดไวยากรณ์จำนวน 143 คำวิจารณ์ รวมทั้งสิ้น 661 คำวิจารณ์

ในการทดสอบความรู้ที่สกัดได้จากระบบ ผลการทดสอบจะพิจารณาใน 2 ประเด็น ได้แก่ ประเด็นที่ 1 การทดสอบความใกล้เคียงของระดับความพึงพอใจที่ระบบสกัดได้กับระดับความพึงพอใจที่นักท่องเที่ยวกำหนด และประเด็นที่ 2 การทดสอบความถูกต้องในการสกัดความรู้แบบอัตโนมัติที่ได้พัฒนาขึ้นมา โดยใช้วิธีวัดค่ามาตรฐานในการประเมินความถูกต้อง 3 ค่า ได้แก่ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) โดยมีสูตรดังสมการที่ 2 3 และ 4 ตามลำดับ (Miao, Duan, Zhang and Jiao, 2009)

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

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

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

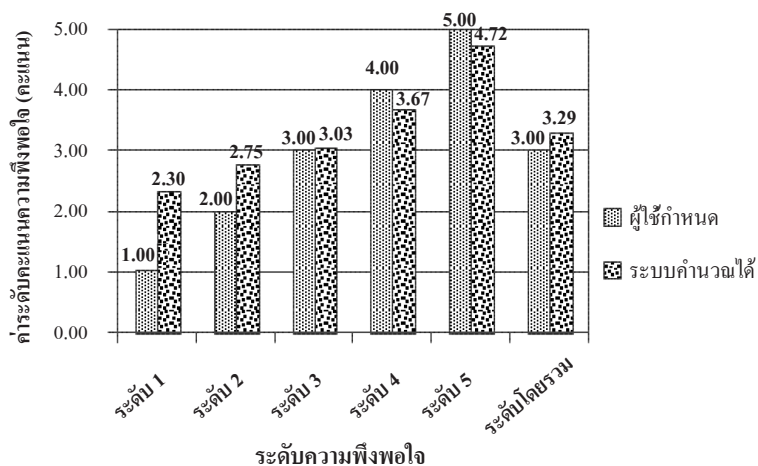
โดย TP (True Positive) หมายถึง ข้อมูลที่สามารถสกัดได้และมีความเกี่ยวข้อง; FP (False Positive) หมายถึง ข้อมูลที่สามารถสกัดได้แต่ไม่มีความเกี่ยวข้อง; TN (True Negative) หมายถึง ข้อมูลที่ไม่สามารถสกัดได้และไม่มีความเกี่ยวข้อง และ FN (False Negative) หมายถึง ข้อมูลที่ไม่สามารถสกัดได้แต่มีความเกี่ยวข้อง

ผลการทดสอบและการอภิปรายผล

จากการจำลองสภาพแวดล้อมเพื่อทดสอบระบบ ผลการทดสอบและการอภิปรายผลในแต่ละแง่มุมมีรายละเอียดดังนี้

ประเด็นที่ 1 การทดสอบความใกล้เคียงของระดับความพึงพอใจที่ระบบสกัดได้กับระดับความพึงพอใจที่นักท่องเที่ยวกำหนด ปรากฏผลการทดสอบดังรูปที่ 6 ซึ่งจากรูปอธิบายได้ว่า เมื่อนำข้อมูลคำวิจารณ์ทั้งหมดมาตรวจสอบในแง่ความใกล้เคียงของระดับความพึงพอใจโดยรวมที่ระบบสกัดได้

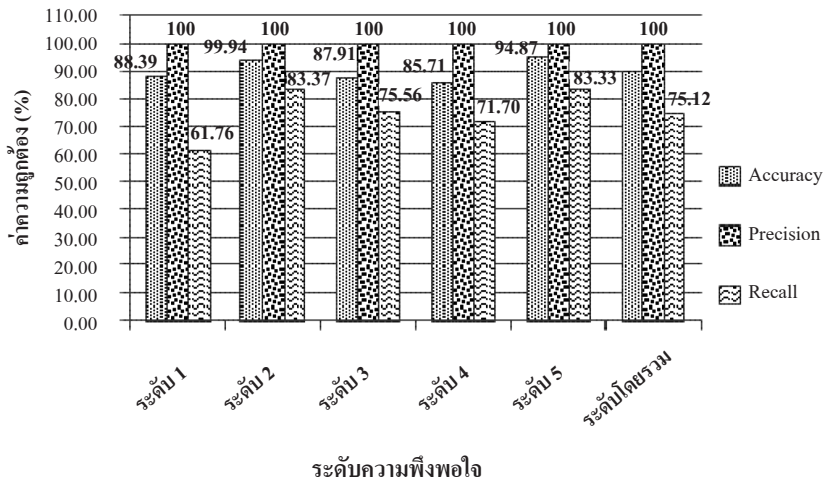
กับระดับความพึงพอใจโดยรวมที่นักท่องเที่ยวกำหนด พบว่ามีความแตกต่างกันจำนวน 0.29 คะแนน (3.29-3.00) ในขณะที่ เมื่อพิจารณาแยกตามกลุ่มคำวิจารณ์พบว่า ระดับความพึงพอใจที่ 3 มีค่าแตกต่างกันน้อยที่สุด (0.03 คะแนน) และระดับความพึงพอใจที่ 1 ซึ่งมีค่าแตกต่างมากที่สุด (1.3 คะแนน) ดังแสดงในรูปที่ 6



ภาพที่ 6 แผนภูมิสรุปผลค่าความแตกต่างของระดับความพึงพอใจที่ระบบสกัดได้
กับระดับความพึงพอใจโดยรวมที่นักท่องเที่ยวกำหนด

โดยเหตุผลที่ทำให้ระดับความพึงพอใจที่ 1 มีค่าแตกต่างกันมากที่สุดเนื่องจาก กลุ่มของกลุ่มไวยากรณ์ไม่พึงบริบทยังไม่ครอบคลุมรูปแบบประโยคคำวิจารณ์ส่วนใหญ่ที่นักท่องเที่ยววิจารณ์ เพราะในบทวิจารณ์ที่แสดงถึงความไม่พอใจในระดับที่ 1 นั้น นักท่องเที่ยวส่วนใหญ่จะใช้ประโยคคำวิจารณ์ในลักษณะบรรยายมากกว่าการกล่าวถึงเรื่องที่่ไม่พอใจโดยตรง ทำให้กระบวนการสกัดความรู้ที่หาที่ประโยคดังที่นำเสนอในงานวิจัยนี้ ไม่สามารถสกัดได้อย่างถูกต้อง เมื่อเปรียบเทียบกับบทวิจารณ์ในระดับความพึงพอใจอื่นๆ ที่นักท่องเที่ยวมักจะวิจารณ์อย่างตรงไปตรงมา การสกัดที่ประโยคจึงมีความถูกต้อง

ประเด็นที่ 2 การทดสอบความถูกต้องในการสกัดความรู้แบบอัตโนมัติที่ได้พัฒนาขึ้นมา ปรากฏผลการทดสอบดังรูปที่ 7 ซึ่งสามารถอภิปรายผลโดยแบ่งเป็น 3 ข้อตามค่ามาตรฐานที่ใช้ในการประเมินความถูกต้องดังที่กล่าวมาแล้ว ดังนี้



ภาพที่ 7 แผนภูมิสรุปผลการประเมินความถูกต้องของการสกัดความรู้

1) ค่าความแม่นยำ (*Precision*) ของทุกระดับความพึงพอใจ พบว่ามีค่าเท่ากับ 100% เสมอ กล่าวคือ ระบบสามารถสกัดข้อมูลที่มีความเกี่ยวข้องกับเรื่องโรงแรมได้ทั้งหมด เนื่องจากการศึกษาครั้งนี้ได้สกัดความรู้โดยใช้ออนโทโลยีที่มีจำนวนคำศัพท์คงที่และมีความเกี่ยวข้องเฉพาะกับเรื่องโรงแรมเท่านั้น ดังนั้น ระบบจะไม่สามารถสกัดข้อมูลที่ไม่มีความเกี่ยวข้องกับเรื่องโรงแรมได้ ($FP = 0$) จึงทำให้ค่าความแม่นยำที่คำนวณได้มีค่าเท่ากับ 100% เสมอ เพราะจากสมการที่ 3 เมื่อแทนค่าจะได้ผลลัพธ์ดังนี้

$$precision = \frac{TP}{(TP + FP)} \times 100\% = \frac{TP}{(TP + 0)} \times 100\% = 100\%$$

2) ค่าความระลึก (*Recall*) ของการสกัดความรู้ พบว่าระดับความพึงพอใจที่ 2 มีค่ามากที่สุด (83.37%) และระดับความพึงพอใจที่ 1 มีค่าน้อยที่สุด (61.76%) โดยมีค่าความระลึกของการสกัดความรู้โดยรวมคิดเป็น 75.12% และสำหรับสาเหตุที่ทำให้ระดับความพึงพอใจที่ 1 มีค่าต่ำ เนื่องจากเหตุผลประการเดียวกับที่กล่าวมาแล้วในประเด็นที่ 1

3) ค่าความถูกต้อง (*Accuracy*) ของการสกัดความรู้ พบว่าระดับความพึงพอใจที่ 5 มีค่ามากที่สุด (94.87%) และระดับความพึงพอใจที่ 4 มีค่าน้อยที่สุด (85.71%) โดยมีค่าความถูกต้องของการสกัดความรู้โดยรวมคิดเป็น 90.15% มีข้อสังเกตว่าระดับความพึงพอใจที่ 1 มีค่าสูงถึง 88.39% ทั้งที่ค่าความระลึกล้นต่ำที่สุด เนื่องจากการวิจารณ์ถึงเรื่องที่ไม่เกี่ยวข้องกับขอบเขตที่ระบบสนใจ

มีอยู่มาก จึงมีผลทำให้ค่าข้อมูลที่ไม่สามารถสกัดได้และไม่มีความเกี่ยวข้อง (TN) สูงกว่า (TN ของระดับ 1 คือ 78 ขณะที่ TN ของระดับ 4 คือ 52) ซึ่งเมื่อพิจารณาจากสมการค่าความถูกต้องคือ

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

พบว่า เมื่อนำค่า TN ดังกล่าวมารวมกับค่า TP จะทำให้ผลรวมที่ได้มีค่าสูงกว่าข้อมูลที่ไม่สามารถสกัดได้แต่มีความเกี่ยวข้อง (FN) เป็นจำนวนมากจนทำให้ค่า FN ส่งผลต่อการคำนวณน้อยหรืออาจไม่ส่งผลเลย นอกจากนี้ ค่า FP ยังมีค่าเท่ากับ 0 เสมอ ดังที่กล่าวมาแล้ว จึงทำให้ค่าความถูกต้องของการสกัดความรู้ที่คำนวณได้ขึ้นอยู่กับค่า TP และ TN เท่านั้น ดังนั้นยัง TN มีค่าสูงก็ยังส่งผลให้ค่าความถูกต้องสูงตามไปด้วย และเมื่อเปรียบเทียบกับค่าความระลึกที่มาจากสมการคือ

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

พบว่า ค่าความถูกต้องจะสูงกว่าค่าความระลึกในทุกระดับความพึงพอใจ เพราะค่าความระลึกไม่มีค่า TN เข้ามาเกี่ยวข้อง

นอกจากนี้ประเด็นปัญหาอื่นๆ ที่ทำให้มีผลต่อความถูกต้องในการสกัดความรู้ สามารถสรุปได้ดังต่อไปนี้

1) ฐานความรู้ยังมีทรัพยากรไม่เพียงพอโดยเฉพาะคำศัพท์ ทั้งในส่วนของคลังคำที่มีความหมายเหมือนกันแต่เขียนต่างกัน คำคุณศัพท์ และคำกริยาวิเศษณ์ ซึ่งทั้งหมดนี้เป็นองค์ประกอบสำคัญที่ทำให้ระบบสามารถสกัดความรู้ได้

2) กลุ่มของกฎไวยากรณ์ในทฤษฎีไวยากรณ์ไม่พึงบริบทที่นำมาใช้ในขั้นตอนการวิเคราะห์เชิงโครงสร้าง ยังไม่ครอบคลุมทุกรูปแบบประโยคคำวิจารณ์ เนื่องจากรูปประโยคคำวิจารณ์ของนักท่องเที่ยวมีเป็นจำนวนมากและมีความหลากหลาย

ทั้งนี้ หากทรัพยากรในฐานความรู้ไม่เพียงพอ หรือกลุ่มของกฎไวยากรณ์ในทฤษฎีไวยากรณ์ไม่พึงบริบทยังไม่ครอบคลุมจะมีผลทำให้ระบบไม่สามารถสกัดข้อมูลที่มีความเกี่ยวข้องได้ ส่งผลให้การคำนวณค่ามาตรฐานในการประเมินความถูกต้องอยู่ในเกณฑ์ค่อนข้างต่ำ และยังทำให้ระดับความพึงพอใจที่ระบบสกัดได้มีค่าน้อยกว่าที่นักท่องเที่ยวกำหนดด้วย

3) การใช้ภาษาที่ไม่ถูกต้องตามหลักไวยากรณ์ภาษาอังกฤษ ทำให้ประโยคคำวิจารณ์ที่ควรนำมาสกัดเป็นความรู้ต้องถูกตัดออกไป จึงทำให้ความรู้หรือระดับความพึงพอใจที่ระบบสกัดได้มีค่าน้อยกว่าความเป็นจริง

4) ระดับคะแนนความพึงพอใจของคำศัพท์ในฐานความรู้ของระบบ เป็นระดับคะแนนที่ได้จากการยืนยันของผู้เชี่ยวชาญด้านภาษา ซึ่งในความเป็นจริงนักท่องเที่ยวบางคนอาจใช้คำศัพท์

ในการแสดงความคิดเห็นที่ไม่สอดคล้องกับระดับคะแนนความพึงพอใจที่ได้จากการยืนยันของผู้เชี่ยวชาญด้านภาษา จึงทำให้ระดับความพึงพอใจที่ระบบสกัดได้มีความแตกต่างจากที่นักท่องเที่ยวกำหนด

บทสรุป

งานวิจัยนี้ได้พัฒนาระบบสกัดความรู้แบบอัตโนมัติจากบทวิจารณ์ออนไลน์ของนักท่องเที่ยวโดยการนำออนโทโลยีมาประยุกต์ใช้ในการจัดเก็บคำศัพท์ เพื่อให้ผลลัพธ์มีความถูกต้องมากที่สุดในการพัฒนาระบบนั้น ได้ออกแบบขั้นตอนการวิเคราะห์และคำนวณระดับความพึงพอใจจากบทวิจารณ์ของนักท่องเที่ยว และประมวลผลการสกัดความรู้ให้อยู่ในรูปแบบที่นักท่องเที่ยวสามารถนำความรู้เหล่านั้นไปใช้ประกอบการตัดสินใจได้ทันที ผลการทดลองพบว่า ระบบที่นำเสนอมีความถูกต้องของการสกัดความรู้เท่ากับ 90.15% ค่าความแม่นยำเท่ากับ 100% ค่าความระลึกเท่ากับ 75.12% และค่าระดับความพึงพอใจของคำวิจารณ์โดยรวมแตกต่างจากระดับความพึงพอใจที่นักท่องเที่ยวกำหนดเท่ากับ 0.29 คะแนน

ทั้งนี้ ระบบสกัดความรู้แบบอัตโนมัติที่นำเสนอจะมีประโยชน์อย่างยิ่ง หากผู้ประกอบการธุรกิจอิเล็กทรอนิกส์นำไปประยุกต์ใช้เก็บและวิเคราะห์ความคิดเห็นของนักท่องเที่ยวเกี่ยวกับสินค้าและบริการ แล้วนำข้อมูลเหล่านั้นมาพัฒนาสินค้าหรือบริการของตนเองให้ตอบสนองความต้องการของนักท่องเที่ยวให้มากที่สุด นอกจากนี้ระบบจะช่วยให้นักท่องเที่ยวสามารถนำความรู้ในเรื่องที่สกัดไปใช้ประกอบการตัดสินใจได้อย่างสะดวกและรวดเร็วอีกด้วย

กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จลุล่วงได้ เนื่องด้วยการสนับสนุนจากทุนอุดหนุนการวิจัยของสำนักงานคณะกรรมการวิจัยแห่งชาติ (วช.) ประจำปีงบประมาณ 2554 อันเป็นส่วนช่วยในการวิจัย เพื่อให้เกิดการสร้างองค์ความรู้ และการพัฒนาผลงานเพื่อนำไปใช้ให้เกิดประโยชน์ต่อสังคมและประเทศชาติต่อไป

เอกสารอ้างอิง

- Chi, C. G-Q., and Qu, H. (2008). Examining the structural relationships of destination image, tourist satisfaction and destination loyalty: An integrated approach. **Tourism Management**, 29(4): 624-636.
- Cracolici, M. F., and Nijkamp, P. (2008). The attractiveness and competitiveness of tourist destinations: A study of Southern Italian regions. **Tourism Management**, 30(3): 336-344.
- Dave, K., Lawrence, S. and Pennock, D.M. (2003). Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. In **Proceeding of the 12th International Conference on World Wide Web** (pp. 519-528). Budapest, Hungary: ACM.

- Del Bosque, I.R., and San Martin, H. (2008). **Tourism Research**. Great Britain: Elsevier. (Annals of Tourism Research. 35(2): 551–573)
- Fry, J., (2004). **Context-Free Grammars for English** [Powerpoint]. California, United States: San José State University.
- Hsu, T-K., Tsai, Y-F., and Wu, H-H. (2009). The preference analysis for tourist choice of destination: A case study of Taiwan. **Tourism Management**. 30(2): 288–297.
- Hu, M. and Liu, B. (2004). Mining and Summarizing Customer Reviews. In **Proceeding of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining** (pp. 168-177). Washington, USA: ACM.
- Miao, D., Duan, Q., Zhang, H., and Jiao, N. (2009). Rough set based hybrid algorithm for text classification. **Expert Systems with Applications**. 36(5): 9168–9174.
- Miniwatts, M.G. (2010). **World Internet Usage and Population Statistics** [On-line]. Available: <http://www.internetworldstats.com/stats3.htm>
- Saaty, T.L. (2006). **Fundamentals of decision making and priority theory with the analytic hierarchy process** (2nd ed.). Pittsburgh, PA: RWS.
- Siorpaes, K., Prantner, K., and Bachlechner, D. (2004). **Class Hierarchy for the e-Tourism Ontology Version 8** [PDF]. Galway, Ireland: Digital Enterprise Research Institute.
- Taboada, M. and Grieve, J. (2004). Analyzing Appraisal Automatically. In **Proceeding of the AAAI Spring Symposium**. (pp. 158-161). Menlo Park, California: AAAI.
- Turney, P.D., (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In **Proceeding of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)** (pp. 417-424). Philadelphia, America: Association for Computational Linguistics.
- Whitman, B. and Ellis, D. (2004). Automatic record reviews. In **Proceeding of the 5th International Society for Music Information Retrieval** (pp. 470-477). Barcelona, Spain: Universitat Pompeu Fabra.
- World Tourism Organization. (2007). **Yearbook of Tourism Statistics**. Madrid: WTO. (Yearbook of Tourism Statistics: Data 2001-2005 the 59th Edition)
- Zheng, W. and Ye, Q. (2009). Sentiment Classification of Chinese Traveler Reviews by Support Vector Machine Algorithm. In **Proceeding of 2009 Third International Symposium on Intelligent Information Technology Application** (pp. 335-338). NanChang, China. IEEE.