

การพัฒนาระบบสรุปทวิจารย์สินค้าภาษาไทย โดยผู้บริโภคมอบอัตโนมัติ

ตรีเทพ ชำรงลักษณะ*
ดร.จันทร์เจ้า มงคลนาวิน**

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์หลักเพื่อพัฒนาระบบสรุปทวิจารย์สินค้าภาษาไทยโดยผู้บริโภคมอบอัตโนมัติ ซึ่งจะสรุปทวิจารย์ของผู้บริโภคที่ได้แสดงความคิดเห็นไว้บนเว็บไซต์พาณิชย์อิเล็กทรอนิกส์ โดยจะสรุป



คุณลักษณะของสินค้าที่ผู้บริโภคให้ความสำคัญและทิศทางความเห็นที่ผู้บริโภคมอบต่อคุณลักษณะนั้นๆ ของสินค้า ว่าเป็นความคิดเห็นในเชิงบวกหรือเชิงลบ ในงานวิจัยนี้ได้มีการประยุกต์ใช้เทคนิคต่างๆ ไม่ว่าจะเป็นการตัดคำ การกำหนดหน้าที่ของคำ การสร้างและกำหนดทิศทางในรายการคำแสดงความคิดเห็นต้นกำเนิด การแยกคำแสดงคุณลักษณะของสินค้า และการระบุทิศทางของคำแสดงคุณลักษณะ มาเป็นส่วนประกอบของระบบสรุปทวิจารย์สินค้า รวมถึงศึกษาค่าคงที่ต่างๆ ที่เหมาะสมกับทวิจารย์สินค้าภาษาไทย ในการทดลอง ผู้วิจัยใช้ข้อมูลทวิจารย์สินค้าประเภทเครื่องสำอาง 4 ชนิด จากเว็บไซต์จีบัน (www.jeban.com) ซึ่งเป็นเว็บไซต์สำหรับแลกเปลี่ยนความคิดเห็นเกี่ยวกับสินค้าเครื่องสำอางและเปรียบเทียบผลสรุปทวิจารย์สินค้าที่ได้จากระบบกับผลสรุปทวิจารย์ชุดเดียวกันจากผู้เชี่ยวชาญทางด้านการตลาด 3 ท่าน ผลปรากฏว่าระบบต้นแบบมีความแม่นยำในการระบุคุณลักษณะและทิศทางของความคิดเห็นในระดับพอใช้ โดยผู้วิจัยได้เสนอแนะเกี่ยวกับบริบทที่เหมาะสมสำหรับการนำระบบต้นแบบไปประยุกต์ใช้งาน

คำสำคัญ : การสรุปโดยคุณลักษณะ การจำแนกคุณลักษณะ ทวิจารย์สินค้าโดยผู้บริโภค การจัดกลุ่มความคิดเห็น

* นิสิตในหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาการพัฒนาซอฟต์แวร์ด้านธุรกิจ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

** ผู้ช่วยศาสตราจารย์ประจำภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

Abstract

The main objective of this study is to develop an automatic system for summarizing Thai customer reviews which would summarize customer reviews posted on e-commerce websites. The system would extract features of product that customers pay attention and summarize whether opinion that they have towards the features is positive or negative. In this study, we applied techniques such as word segmentation, part-of-speech tagging, creation of seedlist of opinion words, feature extraction, and feature orientation identification in different parts of the system. The study is also aimed at discovering related system configurations that suit customer reviews written in Thai. We conducted the experiment on four groups of cosmetic product reviews from www.jeban.com which is a website for sharing customer opinion towards cosmetic products and compared the result from the system against summary of the same set of customer reviews obtained from three marketing experts. The result shows that the accuracy of the prototype system is acceptable. The contexts that are appropriate for applying the system are also suggested.



Keywords : Feature based summarization, Feature extraction, Customer reviews, Sentiment classification

1. บทนำ

ในยุคที่ธุรกิจพาณิชย์อิเล็กทรอนิกส์เติบโตขึ้นอย่างรวดเร็วอย่างเช่นในปัจจุบัน เว็บไซต์พาณิชย์อิเล็กทรอนิกส์ส่วนใหญ่ได้เปิดช่องทางให้ผู้บริโภคแสดงความคิดเห็นเกี่ยวกับสินค้าและบริการของเว็บไซต์เพื่อที่จะเพิ่มความน่าเชื่อถือในตัวสินค้าและทราบถึงความพอใจของผู้บริโภคจากประสบการณ์การใช้สินค้า ผู้บริโภคส่วนใหญ่นิยมที่จะอ่านบทวิจารณ์สินค้าในเว็บไซต์ก่อนที่จะตัดสินใจซื้อ ผลการสำรวจผู้ซื้อสินค้าออนไลน์ (Werbler and Harris, 2008) พบว่า 61 เปอร์เซ็นต์ของจำนวนผู้ซื้อทั้งหมดจะหาบทวิจารณ์สินค้าอ่านก่อนที่จะตัดสินใจเลือกซื้อสินค้าและบริการ แต่ปัญหาที่พบคือ จำนวนบทวิจารณ์สินค้าโดยผู้บริโภคที่เพิ่มขึ้นอย่างรวดเร็วและต่อเนื่อง โดยสินค้าที่เป็นที่นิยมอาจมีจำนวนบทวิจารณ์มากถึง 1,000 บทวิจารณ์ จึงเป็นการยากที่ผู้บริโภคจะเข้าถึงบทวิจารณ์ทั้งหมดและสรุปข้อมูลที่ต้องการเพื่อนำไปประกอบการตัดสินใจ นอกจากนี้ การอ่านข้อมูลแค่บางส่วนอาจทำให้ได้รับข้อสรุปที่ผิดพลาดได้นอกจากผู้บริโภคแล้ว สรุปบทวิจารณ์ดังกล่าวยังเป็นที่ต้องการสำหรับผู้ประกอบการเพื่อที่จะนำมาวิเคราะห์และปรับปรุงคุณภาพของสินค้าให้ตรงกับความต้องการของผู้บริโภค เพราะฉะนั้นการพัฒนาระบบสรุปบทวิจารณ์สินค้าโดยผู้บริโภคแบบอัตโนมัติจึงเป็นสิ่งที่เป็นประโยชน์ต่อทั้งผู้บริโภคและผู้ประกอบการเพื่อให้ได้ข้อมูลที่สะท้อนความคิดเห็นของผู้บริโภคที่มีต่อสินค้าในเวลาที่รวดเร็ว

การพัฒนาระบบสรุปบทวิจารณ์สินค้าโดยผู้บริโภคแบบอัตโนมัติได้มีการศึกษาโดยนักวิจัยบางกลุ่มมาแล้วบ้างในอดีต อย่างไรก็ตามระบบที่ได้มีการศึกษามาก่อนหน้านี้ได้รับการพัฒนาให้มีความเหมาะสมกับลักษณะของภาษาอังกฤษเป็นหลัก ผู้วิจัยจึงมีความสนใจที่จะศึกษาและพัฒนาระบบสรุปบทวิจารณ์สินค้าภาษาไทย เพื่อให้ทราบถึงการกำหนดค่าเปอร์เซ็นต์ต่ำสุด ค่าปัจจัยการจำแนก

และค่าระยะช่องคำที่เหมาะสมสำหรับระบบและบริบทที่เหมาะสมกับการนำไปใช้งาน

2. วรรณกรรมปริทัศน์

ในส่วนต่อไปนี้จะเป็นการทบทวนงานวิจัยที่ได้มีการศึกษามาก่อนในอดีตและงานวิจัยอื่นๆ ที่เกี่ยวข้องกับองค์ประกอบของระบบดังนี้

เนื่องจากภาษามนุษย์มีความซับซ้อน การพัฒนาระบบสรุปบทวิจารณ์สินค้าแบบอัตโนมัติ จึงจำเป็นต้องใช้องค์ประกอบด้านเทคนิคในหลายส่วน ดังเช่น ในงานวิจัยของ Hu และ Liu (Hu and Liu, 2004) ซึ่งได้ใช้เทคนิคการแบ่งประเภทความคิดเห็นในบทวิจารณ์โดยใช้วิธีเวิร์ดเน็ต (Wordnet) มาช่วยในการค้นหาความสัมพันธ์ของคำ เพื่อแบ่งแยกทิศทางของคำแสดงความคิดเห็น ออกเป็นความคิดเห็นเชิงบวกและความคิดเห็นเชิงลบ ในส่วนการค้นหาการแสดงคุณลักษณะของสินค้า Hu และ Liu ได้ใช้วิธีการทำเหมืองข้อมูลแบบหาความสัมพันธ์ (Association Mining) ในการหาเซตความถี่ของคำ ต่างจากงานวิจัยของ Popescu และ Etzioni (Popescu and Etzioni, 2005) ที่ได้คิดค้นวิธี โอพีโอเอ็นอี (OPINE) ซึ่งเป็นการดึงคุณลักษณะของสินค้าและความคิดเห็นในบทวิจารณ์สินค้า โดยนำระบบการดึงข้อมูลแบบเป็นอิสระจากชุดข้อมูล (Domain-independent Information Extraction System) ที่เรียกว่าระบบโนวิทอ (KNOWITALL) (Etzioni et al., 2005) มาใช้ ซึ่งนอกจากจะดึงคุณลักษณะของสินค้าและส่วนประกอบของสินค้าแล้ว ระบบยังสามารถหาคุณลักษณะที่มีความหมายเป็นนัย (Implicit Feature) เช่นคำว่า “tiny” กับคำว่า “small” ซึ่งคำว่า “tiny” มีความหมายใกล้เคียงกับคำว่า “small” สามารถใช้บอกความหมายเป็นนัยได้ว่า “เล็ก” เหมือนกัน นอกจากนี้ระบบดังกล่าว ยังสามารถระบุทิศทางของคำแสดงความคิดเห็นได้จากการหาความสัมพันธ์ของคำแสดงความคิดเห็น

ที่มีต่อคุณลักษณะในบทวิจารณ์ ส่วนงานวิจัยของ Oelke et al. (Oelke et al., 2009) ได้นำวิธีการหาคุณลักษณะของสินค้าโดยใช้วิธีการดึงคำสำคัญ (Key Word) ซึ่งใช้กันมากในการค้นคืนข้อมูล (Information Retrieval) โดยนำมาตราวัดค่าคะแนนที่เอฟไอดีเอฟ (TFIDF) ซึ่งเป็นค่าคะแนนความสำคัญของคำในเอกสาร มาปรับใช้กับคำที่ปรากฏในบทวิจารณ์ของสินค้า (หรือคลาส) ที่มีมากกว่า 1 สินค้า เรียกว่า มาตราวัดค่าคะแนนที่เอฟไอซีเอฟ (TFICF) โดยการนำคำที่เอฟไอซีเอฟของคำในบทวิจารณ์ของแต่ละสินค้ามาเปรียบเทียบกัน ถ้าคำในสินค้าใดมีค่าคะแนนมากกว่าคำเดียวกันนั้นในสินค้าอื่นๆ คำนั้นจะถูกจำแนกออกมาเป็นคุณลักษณะของสินค้านั้นๆ

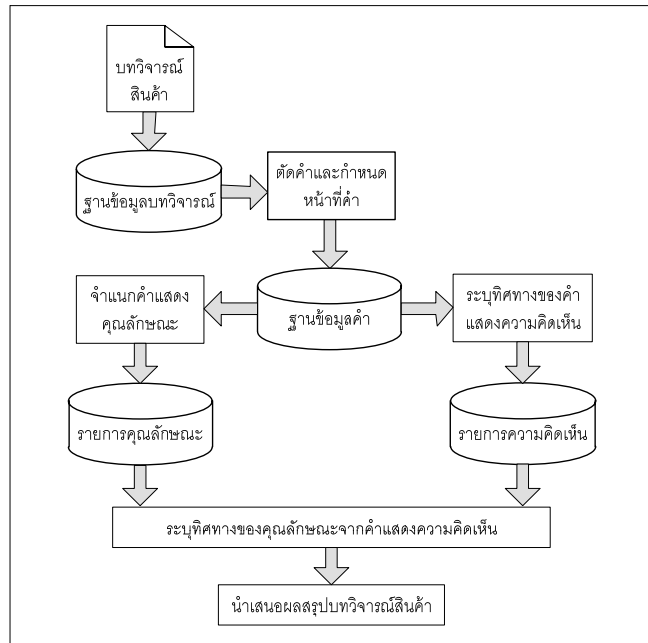
งานวิจัยนี้ได้นำวิธีการของ Oelke et al. (Oelke et al., 2009) มาพิจารณาหาความเหมาะสมของค่าแสดงคุณลักษณะสำหรับใช้ในการสรุปทวิจรณ์สินค้า

ที่มีข้อความภาษาไทย เนื่องจากภาษาไทยเป็นภาษาที่ไม่สามารถหาจุดสิ้นสุดของประโยคได้อย่างชัดเจน จึงทำให้วิธีการหาคุณลักษณะของสินค้าที่มีการประมวลผลแบบประโยคต่อประโยค เช่น วิธีการทำเหมืองข้อมูลแบบเชื่อมโยงกัน (Association Mining) และวิธีโอพีไอเอ็นอี (OPINE) ไม่สามารถนำมาใช้กับบทวิจารณ์ที่เป็นภาษาไทยได้ ในส่วนการหาค่าแสดงความคิดเห็น ผู้วิจัยได้ใช้วิธีการแบ่งประเภทความคิดเห็นด้วยวิธีเวิร์ดเน็ต (Wordnet) เนื่องจากในปัจจุบันยังไม่มีระบบที่ช่วยในการหาความสัมพันธ์ระหว่างคำในภาษาไทยเช่นเดียวกับระบบโนวิทอล (KNOWITALL)

3. องค์ประกอบของระบบ

องค์ประกอบและขั้นตอนของระบบสรุปทวิจรณ์สินค้า โดยผู้บริคแบบอัตโนมัติที่พัฒนาขึ้นในงานวิจัยนี้ สามารถแสดงได้ดังรูปที่ 1

รูปที่ 1 ขั้นตอนของระบบสรุปทวิจรณ์สินค้า



รูปที่ 2 ตัวอย่างบทวิจารณ์สินค้า

สินค้าพร้อมทบทวน

บทวิจารณ์ที่ 1 : ชอบตรงมีเจดีย์ให้เลือกมากมายเข้ากับทุกสีผิวนี้แหละ

บทวิจารณ์ที่ 2 : สีติดทนในระดับหนึ่ง แต่ปิดครั้งแรกไม่ติดค่ะ แต่ครั้งหลังๆ ก็โอเค
แพ็กเกจดูแลยากนิดหนึ่งค่ะ
.....

บทวิจารณ์ที่ n : สีสวย มีให้เลือกเยอะดีค่ะ เม็ดสีชัดเจน ถ้าปิดแรงไประวัง
เหมือนจิว ; p แพคเกจแลดูบอบบาง

สินค้าลองทำความสะอาดหน้าและเครื่องสำอาง

บทวิจารณ์ที่ 1 : ปรับปรุงกลิ่นนิดนึง ใช้แล้วหน้าสะอาดดีค่ะ

บทวิจารณ์ที่ 2 : ก็ล้างเครื่องสำอางออกดีนะค่ะ แต่ต้องล้างหน้าให้เกลี้ยงไม่มันสิ้วตามมาฟรีบ
.....

บทวิจารณ์ที่ n : ใช้แล้วหน้าสะอาดดีค่ะ ผิวชุ่มชื้น กลิ่นไม่หอม แต่เราใช้ยังไม่หมดขวดเลยก็เลิกใช้ซะก่อน

ตารางที่ 1 ตัวอย่างผลสรุปคุณลักษณะและทิศทางความคิดเห็นจากระบบ

สินค้า	คุณลักษณะ	เชิงบวก	เชิงลบ	ผลต่าง	สรุปทิศทาง
สินค้าที่ 1	คุณลักษณะที่ 1	<จำนวน>	<จำนวน>	<จำนวน>	+/-
	คุณลักษณะที่ 2	<จำนวน>	<จำนวน>	<จำนวน>	+/-
	คุณลักษณะที่ 3	<จำนวน>	<จำนวน>	<จำนวน>	+/-
สินค้าที่ 2	คุณลักษณะที่ 1	<จำนวน>	<จำนวน>	<จำนวน>	+/-
	คุณลักษณะที่ 2	<จำนวน>	<จำนวน>	<จำนวน>	+/-

ระบบสรุปบทวิจารณ์สินค้าที่พัฒนาขึ้น มีการนำบทวิจารณ์สินค้าแบบมีโครงสร้าง (ตัวอย่างดังรูปที่ 2) เป็นข้อมูลนำเข้าระบบ (Input) กล่าวคือ ต้องมีการระบุโครงสร้างชัดเจนว่า บทวิจารณ์แต่ละบทวิจารณ์เริ่มต้นและสิ้นสุดเมื่อใด และเป็นบทวิจารณ์ของสินค้าใด โดยระบบที่พัฒนาขึ้นไม่รองรับข้อมูลแบบไม่มีโครงสร้าง (Free Text) ในส่วนของผลลัพธ์ผู้วิจัยได้ออกแบบให้

ระบบมีผลลัพธ์ดังแสดงในตารางที่ 1 เพื่อให้ผู้ใช้ระบบสามารถทราบถึงคุณลักษณะของสินค้าที่ผู้เขียนบทวิจารณ์ให้ความสำคัญและทราบถึงสรุปทิศทางความคิดเห็นที่ผู้เขียนบทวิจารณ์มีต่อคุณลักษณะแต่ละคุณลักษณะได้โดยง่าย ภายในระบบจะประกอบไปด้วย 5 ขั้นตอน ได้แก่ (1) การตัดแบ่งคำในบทวิจารณ์สินค้าและกำหนดหน้าที่คำ (2) การสร้างรายการคำแสดงความคิดเห็นและ

ระบุทิศทางให้กับคำแสดงความคิดเห็น (3) การแยกคำแสดงคุณลักษณะเด่นของสินค้าแต่ละตัวออกจากรายการคำที่ได้จากขั้นตอนที่ 1 (4) การค้นหาคำแสดงความคิดเห็น (ผลลัพธ์จากขั้นตอนที่ 2) ที่มีต่อคุณลักษณะ (ผลลัพธ์จากขั้นตอนที่ 3) แต่ละคุณลักษณะเพื่อนำมาใช้กำหนดทิศทางความคิดเห็นที่ผู้บริโภคมีในแต่ละครั้งที่กล่าวถึงคุณลักษณะนั้นๆ (5) การสรุปความคิดเห็นที่ผู้บริโภคโดยรวมมีต่อคุณลักษณะของสินค้าแต่ละด้านโดยรายละเอียดของแต่ละขั้นตอนมีดังต่อไปนี้

3.1 การตัดคำและกำหนดหน้าที่คำ

เพื่อให้คอมพิวเตอร์สามารถดำเนินการกับคำแต่ละคำได้ บทวิจารณ์สินค้าแต่ละบทวิจารณ์ที่เขียนเป็นข้อความที่ต่อเนื่องกันจำเป็นต้องได้รับการตัดออก

เป็นคำและกำหนดหน้าที่ว่าเป็นคำนาม คำกริยา หรือคำประเภทอื่นๆ ผู้วิจัยจึงนำโปรแกรมตัดคำเล็กส์โต¹ (Thai Lexeme Tokenizer : LexTo) (Haruechaiyasak and Kongyoung, 2009) มาใช้ในการแบ่งคำในบทวิจารณ์และนำคำที่แบ่งได้มากำหนดหน้าที่คำ เพื่อที่จะกรองหน้าที่คำที่ไม่ใช่ในระบบออก ซึ่งได้แก่ คำช่วยกริยา, คำสรรพนาม, คำลักษณนาม, คำบอกกำหนด, คำอุทาน และคำแสดงการถาม เพื่อลดจำนวนคำที่ไม่เกี่ยวข้องกับงานวิจัยและนำเฉพาะคำที่เกี่ยวข้องกับงานวิจัย ซึ่งได้แก่ คำนาม, คำกริยา, คำคุณศัพท์, คำกริยาวิเศษณ์, คำนิเสธ, คำเชื่อม และคำลงท้าย นำมาเก็บในฐานข้อมูล โดยแต่ละหน้าที่คำจะมีการนำไปใช้ในส่วนต่างๆ ของระบบ ดังแสดงในตารางที่ 2

ตารางที่ 2 แสดงหน้าที่คำที่ใช้ในระบบสรุปบทวิจารณ์สินค้าโดยผู้บริโภค

หน้าที่คำ	ลักษณะการนำไปใช้
คำนาม (Noun)	ระบุคุณลักษณะ
คำกริยา (Verb)	ระบุคุณลักษณะ,แสดงความคิดเห็น
คำคุณศัพท์ (Adjective)	แสดงความคิดเห็น
คำกริยาวิเศษณ์ (Adverb)	แสดงความคิดเห็น
คำนิเสธ (Negation)	กลับทิศทางความหมาย
คำเชื่อม (Conjunction)	ระบุจุดสิ้นสุดประโยค
คำลงท้าย (Ending)	ระบุจุดสิ้นสุดประโยค

¹ โปรแกรมตัดคำเล็กส์โต เป็นโปรแกรมตัดคำที่ได้รับรางวัลชนะเลิศจากการแข่งขันซอฟต์แวร์แบ่งคำภาษาไทย (Benchmark for Enhancing the Standard of Thai Language Processing: BEST 2009) ผู้วิจัยได้ทดลองเปรียบเทียบความถูกต้องของโปรแกรมตัดคำเล็กส์โต กับโปรแกรมตัดคำสวอท (Smart Word Analysis for Thai : SWATH) (Charoenpornasawat, 1999; Meknavin et al., 1997) ที่ได้รับการใช้งานอย่างแพร่หลายในช่วงเวลาที่ผ่านมา บนข้อมูลบทวิจารณ์ที่ใช้ในการทดลอง และพบว่าโปรแกรมตัดคำเล็กส์โตมีความถูกต้องในการตัดคำสูงกว่า

บทวิจารณ์สินค้าที่ผ่านการตัดคำแล้วแต่ละบทวิจารณ์จะถูกนำมารวมกัน โดยมีการแบ่งลำดับบทวิจารณ์ที่นำเข้ามาแยกตามสินค้า มีการใช้เครื่องหมาย “|” แบ่งคำออกจากกันและใช้เครื่องหมาย “<ตัวเลข>”

ในการระบุลำดับเลขบทวิจารณ์และใช้เครื่องหมาย “<End>” เป็นตัวบ่งจุดสิ้นสุดของบทวิจารณ์นั้น โดยมีตัวอย่างดังแสดงในรูปที่ 3

รูปที่ 3 ตัวอย่างบทวิจารณ์ที่ผ่านการตัดคำแล้ว

<1>เราว่ากลิ่นก็โอเคนะคะเราไม่ได้ใช้ล้างเครื่องสำอางเท่าไรส่วนมากใช้ในวดหน้าคะ | เวิร์กนะใช้แล้วหน้านุ่ม | นวดได้กลิ่นมือง่ายดี<End><2>|ปรับปรุงกลิ่นนิดนึง | ใช้แล้วสะอาดหน้าดีคะ | แต่เป็นกังวลก็ตอนกลางวันล้างไม่สะอาดเนี่ยะ| เหละ| เลยต้องตามด้วยโฟมล้างหน้าตามอีกครั้ง|<End>

3.2 การสร้างรายการคำแสดงความคิดเห็นและการกำหนดทิศทาง

ในตารางที่ 2 คำที่มีหน้าที่คำเป็นคำคุณศัพท์ คำกริยา และคำกริยาวิเศษณ์ ที่นำมาจากบทวิจารณ์สินค้า จะถูกใช้เป็นคำแสดงความคิดเห็น ซึ่งคำเหล่านี้ใช้ขยายคำที่อยู่ในกลุ่มหน้าที่คำที่จะนำไปใช้ระบุคุณลักษณะ โดยคำคุณศัพท์และคำกริยาจะขยายคำนาม และคำกริยาวิเศษณ์จะขยายคำกริยา ขั้นตอนนี้เป็นขั้นตอนของการสร้างรายการคำแสดงความคิดเห็นและจำแนกกลุ่มคำแสดงความคิดเห็นเหล่านั้นออกเป็นกลุ่มคำแสดงความคิดเห็นเชิงบวกและเชิงลบ

ในการระบุทิศทางให้กับคำแสดงความคิดเห็นนั้นต้องมีรายการคำแสดงความคิดเห็นที่มีการระบุทิศทางไว้แล้วหรือรายการต้นกำเนิด (Seedlist) เป็นรายการเริ่มต้นสำหรับใช้ระบุทิศทางให้กับคำใหม่ที่จะนำเข้ามาเพิ่มในรายการ โดยผู้วิจัยได้ใช้วิธีเวิร์ดเน็ต (Wordnet) (Fellbaum, 1998) ในการค้นหาความสัมพันธ์ของคำระหว่างคำในรายการต้นกำเนิดกับคำใหม่ที่จะนำเข้ามา

กล่าวคือ หากคำใหม่มีความสัมพันธ์แบบพ้องความหมาย (Synonym) กับคำในรายการต้นกำเนิด ระบบจะระบุทิศทางให้กับคำใหม่เป็นทิศทางเดียวกันกับทิศทางของคำในรายการต้นกำเนิด แต่หากคำใหม่มีความสัมพันธ์แบบตรงข้ามความหมาย (Antonym) กับคำในรายการต้นกำเนิด ระบบจะระบุทิศทางให้กับคำใหม่เป็นทิศทางตรงข้ามกับทิศทางของคำนั้นในรายการต้นกำเนิด ในกรณีที่คำใหม่ไม่สามารถหาความสัมพันธ์กับคำในรายการต้นกำเนิดได้ ระบบจะเก็บคำๆ นั้นไว้หาความสัมพันธ์กับคำในรายการต้นกำเนิดอีกครั้งหลังจากเสร็จสิ้นการระบุทิศทางของคำแสดงความคิดเห็นทุกคำในรอบแรกแล้วในรอบที่สอง ระบบจะนำคำแสดงความคิดเห็นที่ถูกเก็บไว้ในรอบแรกมาหาความสัมพันธ์กับคำในรายการต้นกำเนิดที่เป็นผลลัพธ์ที่ได้จากรอบแรก โดยระบบจะทำซ้ำเช่นนี้จนกระทั่งไม่พบความสัมพันธ์ระหว่างคำในรายการคำแสดงความคิดเห็นที่ถูกเก็บไว้กับคำใดๆ ในรายการต้นกำเนิด รายการคำแสดงความคิดเห็นที่ถูกเก็บไว้รอบสุดท้ายจะถูกตัดทิ้งไปจากระบบ โดยมีตัวอย่างของความสัมพันธ์ของคำแสดงความคิดเห็นดังแสดงในรูปที่ 4

รูปที่ 4 ตัวอย่างของความสัมพันธ์ของคำแบบเวิร์ดเน็ต



จากรูปที่ 4 สมมติว่า ในรายการคำต้นกำเนิดมีคำว่า “เร็ว” มีทิศทางเป็นบวก และมีความหมายตรงข้ามกับคำว่า “ช้า” ที่มีทิศทางเป็นลบ หากต้องการเพิ่มคำว่า “ไว” เข้ามาในรายการ ระบบจะหาคำในเวิร์ดเน็ตที่พ้องความหมายกับคำว่า “ไว” หากระบบพบว่า คำว่า “เร็ว” พ้องความหมายกับคำว่า “ไว” และคำว่า “เร็ว” มีทิศทางเป็นบวก คำว่า “ไว” จะถูกระบุให้มีทิศทางเป็นบวก หากต้องการเพิ่มคำว่า “เฉื่อย” ซึ่งพ้องความหมายกับคำว่า “ช้า” ที่มีทิศทางเป็นลบ คำว่า “เฉื่อย” จะถูกระบุให้มีทิศทางเป็นลบในรายการ ในกรณีที่ระบบไม่พบคำพ้องความหมายของคำแสดงความคิดเห็นที่ต้องการเพิ่มใหม่ในเวิร์ดเน็ต แต่พบคำตรงข้ามความหมาย คำที่ต้องการเพิ่มใหม่จะถูกกำหนดให้มีทิศทางตรงข้ามกับคำตรงข้ามความหมายที่พบ และดำเนินการเช่นนี้เรื่อยไปจนหมดคำที่มีหน้าที่คำที่จะใช้เป็นคำแสดงความคิดเห็น (ผลลัพธ์จากขั้นตอนที่ 3.2) รายการคำแสดงความคิดเห็นและทิศทางที่ได้จะถูกนำไปใช้ในการระบุทิศทางความคิดเห็นในบทวิจารณ์ที่ผู้บริกอบมีต่อคุณลักษณะของสินค้าแต่ละด้านในขั้นตอนที่ 3.4 ต่อไป

3.3 การสร้างรายการคำแสดงคุณลักษณะ

ในขั้นตอนการจำแนกคำแสดงคุณลักษณะ ระบบจะใช้คำที่มีหน้าที่เป็นคำนามและคำกริยาเป็นคำแสดง

คุณลักษณะ เนื่องจากคุณลักษณะของสินค้าส่วนใหญ่จะปรากฏอยู่ในรูปคำนาม ส่วนคำกริยาโดยมากจะเป็นส่วนขยายคำนาม แต่สามารถสื่อถึงคำนามโดยไม่ต้องมีคำนามปรากฏอยู่ เช่นคำว่า “เกลี่ย” เป็นคำกริยาหมายถึงการเกลี่ยเนื้อครีม ซึ่งมีคำว่า “เนื้อครีม” เป็นคุณลักษณะ แต่ไม่จำเป็นต้องปรากฏพร้อมกันก็สามารถระบุคุณลักษณะของสินค้า หรืออาจจะมี ความหมายในเชิงการใช้งานก็ได้ เช่น ในสินค้าครีมรองพื้นอาจจะมีผู้บริกอบวิจารณ์สินค้าว่า “เกลี่ยง่าย” โดยคำว่า “เกลี่ย” มีหน้าที่คำเป็นคำกริยาแต่สามารถใช้บอกคุณลักษณะของครีมรองพื้นได้ ในงานวิจัยนี้จึงนำคำกริยามาใช้ร่วมกับคำนามเพื่อใช้เป็นคำแสดงคุณลักษณะของสินค้า

สำหรับวิธีที่นำมาใช้ในการดึงคุณลักษณะของสินค้าออกมาจากบทวิจารณ์ ผู้วิจัยได้นำวิธีการดึงคุณลักษณะโดยการจำแนกคำออกจากกลุ่มคำตามงานวิจัยของ Oelke et al. (Oelke et al., 2009) ซึ่งต้องหาค่าคะแนนของคำที่ต้องการนำมาจำแนก ได้แก่ ค่าคะแนนของคำนามและคำกริยาที่ปรากฏในบทวิจารณ์สินค้าในแต่ละสินค้า โดยใช้มาตรวัดที่เอฟไอซีเอฟ (Term Frequency Inverse Class Frequency : TFICF) (Oelke et al., 2009) มาคำนวณหาค่าคะแนนของคำ ซึ่งมีสูตรการคำนวณดังสมการที่ 1

โดยในสมการต่อไปนี้ กำหนดให้ t คือ คำที่สนใจ² (Term) ที่ปรากฏในบทวิจารณ์สินค้า
 C คือ จำนวนคลาสสินค้าทั้งหมดในระบบ (All Class)
 c คือ คลาสสินค้า (Class) เช่น สินค้าโลชั่นทำความสะอาด
 d คือ บทวิจารณ์สินค้า (Reviews)

\forall term $t_i \wedge \forall$ classes c_j with $\in \{1 \dots \#terms\}$ and $\{1 \dots \#classes\}$:

$$weighted - tfidf(t_i, c_j) = distr_weight(t_i, c_j) \cdot tf(t_i, c_j) \cdot icf(t_i) \quad (1)$$

สมการการหาค่าทีเอฟไอซีเอฟ แบบถ่วงน้ำหนัก หรือค่าคะแนนของคำ ประกอบไปด้วยองค์ประกอบที่สำคัญ 3 ค่า ได้แก่ ค่าทีเอฟ (tf), ค่าไอซีเอฟ (icf) และค่าน้ำหนักการแจกแจง (distr_weight) ซึ่งมีรายละเอียดดังนี้

ค่าทีเอฟ (term frequency value : tf) คือ ค่าความถี่ของคำที่ปรากฏ คำนวณได้จาก ผลรวมความถี่ของคำที่สนใจที่ปรากฏในบทวิจารณ์ทั้งหมดของคลาสสินค้าหารด้วย ผลรวมความถี่ของคำทั้งหมดทุกประเภทที่ปรากฏในบทวิจารณ์ทั้งหมด (#d) ของคลาสสินค้านั้น โดยแสดงได้ดังสมการที่ 2

$$tf(t_i, c_j) = \frac{\sum_{k=1}^{\#d_j} freq(t_i, d_{jk})}{\sum_{k=1}^{\#d_j} \#tokens(d_{jk})} \quad (2)$$

ค่าไอซีเอฟ (Inverse Class Frequency : icf) (แสดงดังสมการที่ 3) คือ ค่าความถี่ของคำที่ปรากฏในแต่ละคลาสสินค้า คำนวณได้จาก log ฐานสิบของจำนวนคลาสสินค้าทั้งหมด หารด้วย จำนวนคลาสสินค้าที่มีค่าเปอร์เซ็นต์ความถี่ของคำที่สนใจที่ปรากฏในบทวิจารณ์ของคลาสสินค้านั้น มากกว่าค่าเปอร์เซ็นต์ต่ำสุด (Minimum Percentage) ที่กำหนดไว้ ซึ่งเป็นค่าจำนวนขั้นต่ำสุดของคำที่สนใจที่ปรากฏในบทวิจารณ์ของแต่ละ

คลาสสินค้า เพื่อกรองคำที่ปรากฏอยู่แค่จำนวนไม่กี่คำ ออกจากการพิจารณา โดยผู้ใช้งานสามารถกำหนดค่าเปอร์เซ็นต์ต่ำสุดได้เองตามความเหมาะสมในการใช้งาน เช่น กำหนดให้ค่าเปอร์เซ็นต์ต่ำสุดซึ่งแทนด้วยค่า X มีค่าเป็น 0.2 หมายความว่า ต้องมีคำที่สนใจปรากฏอยู่ในบทวิจารณ์ของคลาสสินค้านั้นเป็นจำนวนอย่างน้อย 20% ของบทวิจารณ์สินค้านั้นทั้งหมด จึงจะสามารถนำคลาสสินค้านั้นมานับรวมเป็นอาหารในสมการที่ 3 ได้

$$icf(t) = \log_{10} \left(\frac{|C|}{|\{c \in C: \frac{|\{d \in C: t \in d\}|}{|d \in c|} > X\}|} \right) \quad (3)$$

² คำนามหรือคำกริยาที่นำมาหาค่าคะแนนของคำ

ค่าน้ำหนักการแจกแจง (Distribution Weight : *distr_weight*) สำหรับการหาค่าน้ำหนักการแจกแจง ผู้วิจัยได้นำวิธีการแจกแจงของคำทั้งหมดในเอกสารหรือบทวิจารณ์ในคลาสสินค้า (Term Relevance Weight) (Salton and Buckley, 1988) มาใช้หาค่าน้ำหนัก ซึ่งสามารถคำนวณได้จากสัดส่วนค่านับสนุน (Support)

ของคลาสสินค้าที่สนใจ ต่อค่านับสนุนของคลาสสินค้าอื่นๆ ดังแสดงในสมการที่ 4 โดยค่านับสนุนหาได้จากจำนวนบทวิจารณ์ในคลาสสินค้าที่มีคำที่สนใจปรากฏหารด้วยจำนวนบทวิจารณ์ทั้งหมดของคลาสสินค้านั้น โดยแสดงได้ดังสมการที่ 5

$$term\ relevance\ weight\ (t_i, c_j) = \frac{support(t_i, c_j)}{\sum_{k \neq j} support(t_i, c_k)} \quad (4)$$

โดย

$$support\ (t_x, c_y) = \frac{| \{d \in c_y : t_x \in d\} |}{| \{d \in c_y\} |} \quad (5)$$

เมื่อกำหนดได้ค่าทีเอฟ (tf), ค่าไอซีเอฟ (icf) และค่าน้ำหนักการแจกแจง (*distr_weight*) แล้ว นำค่าทั้ง 3 มาคูณกันจะได้ค่าทีเอฟไอซีเอฟแบบถ่วงน้ำหนัก (*weighted-tf-icf*) ซึ่งเป็นค่าคะแนนของคำ (Term Score) ที่จะนำมาใช้ในการจำแนกค่าแสดงคุณลักษณะสำหรับแต่ละคลาสจากการเปรียบเทียบค่าคะแนนของคำจากแต่ละคลาสสินค้า โดยจะมีค่าปัจจัยการจำแนก (Discrimination Factor) เป็นตัวกำหนดอัตราการ

จำแนกค่าแสดงคุณลักษณะของแต่ละคลาสสินค้า เพื่อให้คำที่จะใช้เป็นคำแสดงคุณลักษณะของสินค้าหนึ่งๆ เป็นคำที่พบปรากฏในบทวิจารณ์สินค้านั้นๆ มากกว่าที่ปรากฏในบทวิจารณ์สินค้าอื่นๆ อย่างเด่นชัด คำที่ถูกจำแนกออกมาจากแต่ละคลาสสินค้าจะถูกนำไปใช้เป็นคุณลักษณะของสินค้าคลาสนั้นๆ โดยผู้ใช้งานจะเป็นผู้กำหนดค่าปัจจัยการจำแนกเองตามความเหมาะสม ดังแสดงในสมการที่ 6

กำหนดให้ t คือ คำที่จะเลือกเป็นคำแสดงคุณลักษณะ C_k

$$\forall i \in \{1 \dots n\} - k: \quad (6)$$

$$weighted - tficf (t, c_k) > discrimination - factor \cdot weighted - tficf (t, c)$$

3.4 การระบุทิศทางของคุณลักษณะจากความคิดเห็น

ขั้นตอนนี้เป็นขั้นตอนการนำรายการคำแสดงความคิดเห็นและรายการคำแสดงคุณลักษณะมาตรวจหาความคิดเห็นที่มีต่อคุณลักษณะในบทวิจารณ์สินค้าเพื่อสรุปทิศทางความคิดเห็นที่ผู้วิจารณ์โดยรวมมีต่อ

คุณลักษณะแต่ละคุณลักษณะ โดยใช้วิธีการหาค่าน้ำหนักแบบรีเวิร์สดีสแทนซ์ (Reverse-Distance-Weighting: RDW) (Oelke et al., 2009) เป็นวิธีการตรวจหาคำแสดงความคิดเห็นที่มีต่อคุณลักษณะ วิธีการหาค่าน้ำหนักแบบรีเวิร์สดีสแทนซ์ เป็นวิธีที่กำหนดระยะช่องคำ

ในการค้นหาคำแสดงความคิดเห็น โดยหากพบคำแสดงความคิดเห็นในระยะช่องคำที่ระบุไว้แสดงว่าคำๆ นั้น เป็นคำแสดงความคิดเห็นที่มีต่อคุณลักษณะนั้นๆ โดย จะเลือกคำแสดงความคิดเห็นที่อยู่ใกล้กับคุณลักษณะมากที่สุดมาใช้ ดังแสดงในสมการที่ 7

กำหนดให้ $dist$ คือ ระยะระหว่างคำแสดงความคิดเห็นกับคุณลักษณะ
 $cutoff$ คือ ค่าระยะช่องคำ
 f คือ คุณลักษณะของสินค้า
 o คือ คำแสดงความคิดเห็น

$$Rd_weight(f,o) = \begin{cases} 1 & \text{if } dist(f,o) \leq cutoff \\ 0 & \text{else} \end{cases} \quad (7)$$

สมการที่ 7 เป็นการค้นหาตำแหน่งของคำแสดงความคิดเห็น (o) และคุณลักษณะ (f) ในบทวิจารณ์ซึ่งมีการกำหนดค่าระยะช่องคำ (cutoff) เป็นค่าขีดแบ่ง (Threshold) ช่วยป้องกันความผิดพลาดซึ่งเกิดจากระยะทางระหว่างคำแสดงความคิดเห็นกับคุณลักษณะที่ไกลเกินไป ในการตรวจหาคำมีหลักการว่าจะเริ่มหาคำที่อยู่ใกล้คุณลักษณะมากที่สุดทั้งข้างหน้าคำและข้างหลังคำสลับกันไป จนกระทั่งพบคำแสดงความคิดเห็นหรือเมื่อสิ้นสุดระยะช่องคำที่ต้องการตรวจสอบ โดยในงานวิจัยนี้จะเริ่มตรวจจากข้างหน้าคำหรือทางซ้ายมือก่อนแล้วจึงมาตรวจทางขวามือ เช่น กำหนดให้ค่าระยะช่องคำมีค่าเป็น 3 คือ ตรวจหาคำทางซ้ายเป็นระยะ 3 ช่องคำ และทางขวา ระยะ 3 ช่องคำ เริ่มจากทางซ้าย (1) แล้วจึงมาตรวจทางขวา (2) แล้วสลับไปซ้าย (3) และขวา (4) จนครบระยะช่องคำที่กำหนด โดยมีตัวอย่างดังนี้

(5) | (3) | (1) | คุณลักษณะ | (2) | (4) | (6)

หากพบคำแสดงความคิดเห็นในระยะช่องคำที่กำหนดไว้ คำน้ำหนักจะมีค่าเป็น 1 แต่ถ้าไม่พบคำแสดงความคิดเห็นในระยะช่องคำที่กำหนดไว้ คำน้ำหนักจะมีค่าเป็น 0 ซึ่งคำน้ำหนักนี้จะถูกนำไปคำนวณรวมกับค่าทิศทาง (Polarity) ของคำแสดงความคิดเห็นที่มีต่อคุณลักษณะของสินค้าเพื่อใช้ระบุเป็นค่าคะแนนทิศทางให้กับคุณลักษณะ นอกจากนี้ยังมีการหาคำนิเสธ³ (Negation) ที่ปรากฏในประโยคแสดงความคิดเห็น เช่น คำว่า “ไม่” ซึ่งเป็นคำที่ใช้กลับทิศของค่าทิศทางคำแสดงความคิดเห็น สำหรับการตรวจหาคำนิเสธจะใช้วิธีการเดียวกันคือ กำหนดระยะช่องคำในการค้นหา แต่แตกต่างกันตรงที่การค้นหาคำนิเสธจะตรวจเฉพาะทางซ้ายมือเท่านั้น เพราะในภาษาไทยที่ใช้กันทั่วไปในปัจจุบัน คำนิเสธที่มีต่อคำแสดงความคิดเห็นจะปรากฏอยู่ทางด้านหน้าคำเท่านั้น ถ้าตรวจพบคำนิเสธ ค่าทิศทางของคำแสดงความคิดเห็นจะถูกกลับทิศเป็นตรงกันข้ามก่อนนำไปคำนวณค่าคะแนนทิศทางของคุณลักษณะ โดยคำนิเสธ (neg_weight) จะมีคำน้ำหนักเป็น -1 เสมอ ดังแสดงในสมการที่ 8

³ คำนิเสธที่นำมาใช้ในงานวิจัยนี้มาจากคำในฐานข้อมูลคำศัพท์เลือกชิตรอนที่มีหน้าคำเป็นคำนิเสธหรือคำปฏิเสธ ได้แก่ คำว่า “ไม่” “ไม่ใช่” “เช่นนั้น” “เปล่า” “หาไม่” “มิได้” “มิ” “หามิได้” และ “มิใช่”

$$\text{Opinion_score}(f,r) = \sum_{o \in r} \text{rd_weight}(f,o) \cdot \text{polarity}(o) \cdot \text{neg_weight}(n,o) \quad (8)$$

จากสมการที่ 8 ค่าคะแนนทิศทางของคุณลักษณะ (f) ในบทวิจารณ์สินค้า (r) หาได้จากผลรวมค่าน้ำหนักระยะทางของคุณลักษณะ (f) กับความคิดเห็น (o) (rd_weight) คูณกับค่าทิศทางของความคิดเห็น (polarity) คูณกับค่าน้ำหนักระยะทางของคำนิเสธ (n) กับความคิดเห็น (o) (neg_weight)

3.5 การสรุปผลทิศทางความคิดเห็นที่มีต่อคุณลักษณะของสินค้า

หลังจากขั้นตอนที่ 3.4 ที่มีการระบุทิศทางของความคิดเห็นแต่ละความคิดเห็นที่มีต่อคุณลักษณะว่าเป็นความคิดเห็นในทิศทางบวกหรือทิศทางลบ ขั้นตอนนี้จะสรุปคะแนนทิศทางความคิดเห็นที่มีต่อคุณลักษณะแต่ละด้านของสินค้า ในขั้นแรก ค่าคะแนนทิศทางจะได้รับการสรุปรวมเป็น 2 ส่วน คือ ส่วนที่มีค่าคะแนนเป็นบวก (คะแนนของความคิดเห็นเชิงบวก) และส่วนที่มีค่าคะแนนเป็นลบ (คะแนนของความคิดเห็นเชิงลบ) โดยจะมีการคำนวณรวมคะแนนเชิงบวกและเชิงลบ เมื่อรวบรวมค่าคะแนนทิศทางในแต่ละส่วนครบแล้วจึงนำมาหักล้างกันเพื่อสรุปค่าคะแนนทิศทางรวม ถ้าผลลัพธ์เป็นบวก แสดงว่าความคิดเห็นโดยรวมในบทวิจารณ์ที่มีต่อของ

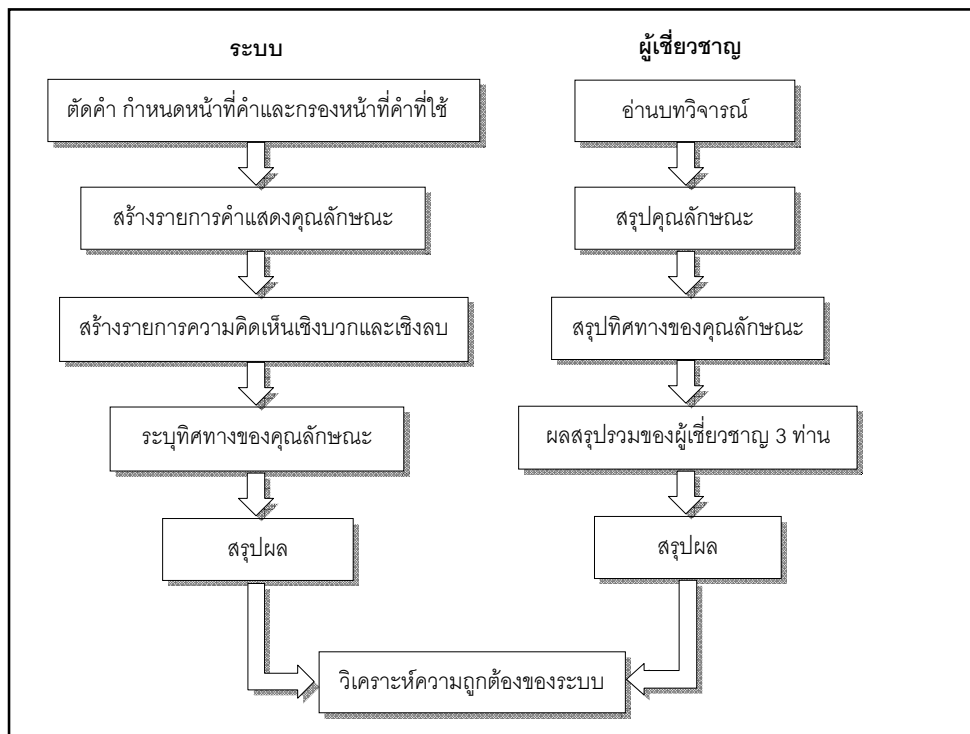
คุณลักษณะนั้นเป็นทิศทางบวก แต่ถ้าผลลัพธ์เป็นลบ แสดงว่าความคิดเห็นโดยรวมในบทวิจารณ์ที่มีต่อคุณลักษณะนั้นเป็นทิศทางลบ

หลังจากได้ผลสรุปค่าคะแนนทิศทางของคุณลักษณะแล้ว จะนำผลลัพธ์มาแสดงผลโดยแสดงชื่อสินค้า, คุณลักษณะที่ได้จากระบบ, ค่าคะแนนทิศทางเชิงบวก, ค่าคะแนนทิศทางเชิงลบ, ผลต่างระหว่างค่าคะแนนเชิงบวกและเชิงลบ และสรุปทิศทางของคุณลักษณะดังแสดงในตารางที่ 1

4. การออกแบบทดลอง

เพื่อศึกษาว่าระบบที่พัฒนาขึ้นมีความถูกต้องในการสรุปบทวิจารณ์สินค้าโดยผู้บริโภคน้อยเพียงใด ผู้วิจัยจึงได้ออกแบบการทดลองโดยแบ่งการทดลองออกเป็น 2 ส่วนได้แก่ ส่วนของระบบกับส่วนของผู้เชี่ยวชาญ ซึ่ง 2 ส่วนนี้จะทำงานแยกกันโดยใช้ข้อมูลเริ่มต้นเป็นบทวิจารณ์สินค้าชุดเดียวกัน แล้วนำผลที่ได้มาเปรียบเทียบกัน โดยมีวัตถุประสงค์เพื่อวัดความถูกต้องของระบบสรุปบทวิจารณ์สินค้าโดยผู้บริโภค โดยมีขั้นตอนดังแสดงในรูปที่ 5

รูปที่ 5 รูปแบบการทดลองในการวัดความถูกต้องของระบบ



4.1 ข้อมูลที่ใช้ในการทดลอง

ข้อมูลที่นำมาใช้ในงานวิจัยนี้ เป็นบทวิจารณ์สินค้าจากเว็บไซต์จีบัน (jeban.com) ซึ่งเป็นเว็บไซต์สำหรับนำเสนอสินค้าประเภทเครื่องสำอางซึ่งมีส่วนแลกเปลี่ยนความคิดเห็นกันระหว่างผู้บริโภค ผู้วิจัยใช้วิธีการสุ่มตัวอย่าง (เว็บไซต์) แบบเจาะจง (Purposive sampling) เนื่องจากเว็บไซต์จีบันเป็นเว็บไซต์ที่มีบทวิจารณ์สินค้าของผู้บริโภคเป็นจำนวนมากเพียงพอที่จะใช้ในการศึกษา กล่าวคือ มีบทวิจารณ์ของสินค้าที่เลือกมาแต่ละชนิดมากกว่า 250

บทวิจารณ์ขึ้นไป ในการเก็บข้อมูลบทวิจารณ์สินค้า ผู้วิจัยใช้วิธีการคัดลอก (Copy) ข้อความในบทวิจารณ์มาเก็บเข้าฐานข้อมูลโดยเก็บที่ละบทวิจารณ์ ในการทดลองจะใช้บทวิจารณ์สินค้า 4 กลุ่มสินค้า⁴ ได้แก่ สินค้ารีซออนทาแก้มชนิดหนึ่ง สินค้าโลชั่นทำความสะอาดหน้าและเครื่องสำอางชนิดหนึ่ง สินค้าครีมรองพื้นชนิดหนึ่ง และสินค้าครีมรองพื้นผสมกันแดดชนิดหนึ่ง ซึ่งมีจำนวนบทวิจารณ์ 800, 300, 300 และ 280 บทวิจารณ์ตามลำดับ

⁴ กลุ่มสินค้าที่นำมาใช้เป็นกลุ่มสินค้าที่มีจำนวนบทวิจารณ์สูงสุด 4 อันดับแรกบนเว็บไซต์จีบัน

4.2 ค่าคงที่ที่ต้องกำหนดในส่วนของระบบ

ขั้นตอนการทดลองในส่วนระบบแบ่งออกเป็น 5 ส่วน ตามขั้นตอนการทำงานของระบบที่ได้กล่าวไปแล้ว ในหัวข้อที่ 3 ซึ่งในแต่ละขั้นตอนจะมีข้อมูลที่ใช้หรือค่าคงที่ที่ต้องกำหนด ดังนี้

4.2.1 รายการต้นกำเนิด

การสร้างรายการต้นกำเนิดเป็นการกำหนดค่าที่ใช้ระบุทิศทางให้กับคำใหม่ที่จะเพิ่มเข้ามาในรายการค่าแสดงความคิดเห็น ในงานวิจัยนี้ได้ทดลองนำรายการความถี่ของคำทั่วไปของโครงการคลังข้อมูลภาษาไทยแห่งชาติ (Thai National Corpus) (Aroonmanakun et al., 2009) มาแยกเอาคำคุณศัพท์และคำกริยาวิเศษณ์ที่จะนำมาใช้เป็นค่าแสดงความคิดเห็นในรายการต้นกำเนิดออกจากรายการคำทั่วไป โดยใช้ฐานข้อมูลเล็กชิตรอนในการระบุหน้าที่ของคำ และนำมาจัดลำดับความถี่สูงสุดของคำ เพื่อที่จะนำคำที่มีความถี่สูงสุด 20 คำแรกมาใช้ ผลปรากฏว่า คำที่ได้ไม่เหมาะสมที่จะนำมาใช้เป็นรายการต้นกำเนิด เนื่องจากความหมายของคำส่วนใหญ่ไม่ได้สื่อถึงความคิดเห็นที่มีต่อคุณลักษณะของสินค้าที่ใช้ในการทดลอง ผู้วิจัยจึงได้เปลี่ยนมาใช้วิธีการสำรวจโดยสร้างแบบสอบถามเพื่อเก็บรวบรวมข้อมูลค่าแสดงความคิดเห็นทั่วไปที่จะใช้วิจารณ์สินค้าแต่ละสินค้าที่ใช้ในการทดลอง ซึ่งผู้ตอบแบบสอบถามในงานวิจัยนี้เป็นนิสิตปริญญาตรีจำนวน 40 คน และนิสิตปริญญาโทจำนวน 180 คน โดยในจำนวนนี้จะมีทั้งผู้ที่มีประสบการณ์และไม่มีความรู้ในการใช้สินค้าในบทวิจารณ์ที่จะใช้ในการทดลอง ผู้ตอบจะเลือกทำแบบสอบถามเฉพาะสินค้าที่ตนเองมีประสบการณ์การใช้งาน แล้วจึงให้ระบุค่าแสดงคุณลักษณะที่ควรจะมีและค่าแสดงความคิดเห็นที่มีต่อคุณลักษณะนั้น โดยแบ่งเป็นความคิดเห็นเชิงบวกและความคิดเห็นเชิงลบ จากนั้นผู้วิจัยก็เก็บรวบรวมเฉพาะค่าแสดงความคิดเห็นโดยแบ่งเป็นเชิงบวกและเชิงลบ แล้วจึงแยกหน้าที่คำ

ออกเป็นคำคุณศัพท์ คำกริยา และคำกริยาวิเศษณ์ โดยจะเลือกคำที่มีความถี่สูงสุด 10 คำแรกในแต่ละหน้าที่คำมาใช้เป็นค่าในรายการต้นกำเนิดและดำเนินการเพิ่มค่าแสดงความคิดเห็นเข้าไปในรายการต้นกำเนิดตามวิธีการที่กล่าวไว้ในหัวข้อ 3.2 โดยมีจำนวนคำต้นกำเนิดที่สรุปได้จากแบบสอบถามและจำนวนคำที่เพิ่มเข้ามาในรายการต้นกำเนิดด้วยวิธีเวิร์ดเน็ต แสดงได้ดังตารางที่ 3

4.2.2 ค่าคงที่ที่ต้องใช้ในการค้นหาคุณลักษณะ

ในขั้นตอนการดึงค่าแสดงคุณลักษณะจากบทวิจารณ์สินค้า มีค่าคงที่ 2 ค่า ที่ต้องมีการทดลองหาค่าที่เหมาะสม ได้แก่ (1) การกำหนดค่าเปอร์เซ็นต์ต่ำสุด (Minimum Percentage) ในการหาค่าไอซีเอฟ (icf) เพื่อนำมาคำนวณหาค่าคะแนนของคำ ซึ่งค่าเปอร์เซ็นต์ต่ำสุดนี้เป็นค่าที่ใช้กรองการกระจายตัวของคำในบทวิจารณ์สินค้าเพื่อป้องกันคำที่มีการกระจายตัวในบทวิจารณ์สินค้าต่ำ ผู้วิจัยจึงได้ทดลองกำหนดค่าเปอร์เซ็นต์ต่ำสุด โดยเริ่มจากค่า 0.1 หรือ 10 เปอร์เซ็นต์ และเพิ่มขึ้นทีละ 0.05 หรือ 5 เปอร์เซ็นต์ สาเหตุที่ไม่เริ่มที่ 0.05 เพราะอาจจะได้ค่าที่ไม่ได้เป็นประเด็นสำคัญในบทวิจารณ์ เนื่องจากจะได้คำที่มีการกระจายตัวน้อยในบทวิจารณ์สินค้ากลุ่มนั้นๆ (2) การกำหนดค่าปัจจัยการจำแนก (Discrimination Factor) เป็นค่าขีดแบ่งในการจำแนกคำ ซึ่งในงานวิจัยของ Oelke et al. (Oelke et al., 2009) ได้ทดลองใช้ค่าปัจจัยการจำแนกระหว่าง 1.5 ถึง 3.0 และได้แนะนำให้ผู้ที่จะนำเทคนิคไปประยุกต์ใช้ปรับค่าปัจจัยการจำแนกตามความเหมาะสม ผู้วิจัยจึงทดลองใช้ค่าปัจจัยการจำแนก โดยเริ่มจากค่าน้อยที่สุดคือ 1 และเพิ่มขึ้นทีละ 0.5

ผลลัพธ์ที่ได้จากการทดลองกำหนดค่าเปอร์เซ็นต์ต่ำสุดและค่าปัจจัยการจำแนกพบว่า ค่าเปอร์เซ็นต์ต่ำสุดที่เหมาะสมที่สุดคือ 0.1 เนื่องจากได้ค่าที่เป็นผลลัพธ์มากกว่าค่าอื่นๆ และค่าที่เพิ่มขึ้นมานั้นเป็นค่าที่มีความหมายที่สามารถบ่งบอกความเป็นคุณลักษณะของสินค้าได้มากกว่าค่าอื่นๆ โดยได้ทดลองกำหนดค่าเปอร์เซ็นต์

ตารางที่ 3 แสดงจำนวนคำในรายการต้นกำเนิดและจำนวนคำที่เพิ่มเข้ามาในรายการต้นกำเนิด

สินค้า	หน้าที่คำในรายการต้นกำเนิด	จำนวนคำในรายการต้นกำเนิด	จำนวนคำที่เพิ่มเข้ามาในรายการต้นกำเนิดด้วยวิธีเวิร์ดเน็ต
บริษัทออนทาแก้ม	คำคุณศัพท์	19	27
	คำกริยาวิเศษณ์	9	10
	คำกริยา	18	15
โลชั่นทำความสะอาดหน้าและเครื่องสำอาง	คำคุณศัพท์	20	23
	คำกริยาวิเศษณ์	15	14
	คำกริยา	20	19
ครีมรองพื้น	คำคุณศัพท์	15	19
	คำกริยาวิเศษณ์	4	5
	คำกริยา	14	13
ครีมรองพื้นผสมกันแดด	คำคุณศัพท์	19	24
	คำกริยาวิเศษณ์	11	9
	คำกริยา	20	18

ต่ำสุดระหว่าง 0.1 ถึง 0.2 ส่วนคำปัจจัยการจำแนกที่เหมาะสมที่สุดคือ 1.5 เนื่องจากคำที่ได้นั้นมีความเหมาะสมที่สุด คือ เป็นคำที่บ่งบอกถึงส่วนประกอบที่เป็นคุณลักษณะของสินค้าได้ชัดเจน และมีปริมาณคำพอเหมาะ โดยได้ทดลองกำหนดคำปัจจัยการจำแนกระหว่าง 1 ถึง 2 โดยหากเพิ่มคำปัจจัยการจำแนกเป็น 2 แล้ว คำที่ได้จะมีปริมาณน้อย แต่ถ้ากำหนดคำปัจจัยการจำแนกเป็น 1 คำ ที่ได้เพิ่มขึ้นมาส่วนใหญ่เป็นคำที่ไม่เหมาะสมที่จะนำมาเป็นคุณลักษณะของสินค้า คือ เป็นคำที่ไม่ได้เกี่ยวข้องกับส่วนประกอบที่เป็นคุณลักษณะของสินค้า เช่น คำว่า “กก” “อยู่” “คือ” ฯ

4.2.3 ค่าขอบเขตที่ใช้ในการตรวจหาความคิดเห็นและคำนิเสธ

เป็นการทดลองกำหนดค่าระยะช่องคำ (cutoff) ในการตรวจหาคำแสดงความคิดเห็นที่มีต่อคุณลักษณะ และค่าระยะช่องคำในการตรวจหาคำนิเสธที่มีต่อคำแสดงความคิดเห็น ผู้วิจัยได้ทดลองกำหนดค่าระยะช่องคำ

จาก 2 ช่องคำ และเพิ่มขึ้นทีละ 1 ช่องคำ โดยเปรียบเทียบผลลัพธ์ที่ได้ในแต่ละค่าระยะช่องคำ ว่าคำแสดงความคิดเห็นที่มีต่อคุณลักษณะที่พบมีความเหมาะสมมากขึ้นหรือน้อยลง และค่าใดที่มีความเหมาะสมมากที่สุด โดยพิจารณาจากจำนวนคำแสดงความคิดเห็นที่พบเพิ่มขึ้นและจำนวนคำแสดงความคิดเห็นที่มีความสัมพันธ์กับคุณลักษณะที่เพิ่มขึ้น

จากการทดลองพบว่า ค่าระยะช่องคำในการตรวจหาคำแสดงความคิดเห็นที่มีต่อคุณลักษณะที่เหมาะสมคือ ค่าระยะช่องคำที่ 3 เนื่องจากที่ค่าระยะช่องคำที่ 3 จะได้คำแสดงความคิดเห็นที่สัมพันธ์กับคุณลักษณะเพิ่มขึ้นมากที่สุด และได้คำแสดงความคิดเห็นที่ไม่สัมพันธ์กับคุณลักษณะน้อย นอกจากนี้ จากการวิเคราะห์รูปแบบประโยคแสดงความคิดเห็นพบว่า ประโยคแสดงความคิดเห็นที่พบได้บ่อยจะมีระยะห่างของความคิดเห็นจากคุณลักษณะในช่วง 1-3 ช่องคำ เช่น



ส่วนผลการทดลองคำระยะช่องคำในการตรวจหาคำนิเสธที่มีต่อคำแสดงความคิดเห็นพบว่า คำที่เหมาะสมที่สุดคือ คำระยะช่องคำที่ 3 เช่นเดียวกัน โดยพิจารณาจากความถูกต้องและความผิดพลาดที่เกิดขึ้น ซึ่งพบว่าที่ระยะ 3 ช่องคำ พบประโยคที่มีคำนิเสธปรากฏอยู่ในประโยคความคิดเห็นที่ต้องเพิ่มขึ้นจากระยะ 2 ช่องคำ มา 9 ประโยค แต่พบประโยคที่ผิดพลาดคือคำนิเสธไม่ได้ขยายคำแสดงความคิดเห็นในประโยคเดียวกันเพิ่มขึ้นมา 3 ประโยค แต่เมื่อเพิ่มคำระยะช่องเป็น 4 กลับพบประโยคที่ต้องเพิ่มขึ้น 5 ประโยคและประโยคที่ผิดพลาดเพิ่มขึ้น 4 ประโยค เมื่อเปรียบเทียบความถูกต้องและความผิดพลาดที่เพิ่มขึ้นแล้ว ผู้วิจัยเห็นว่าที่คำระยะช่องคำที่ 3 มีความถูกต้องที่เพิ่มขึ้นมากกว่าความผิดพลาดที่เกิดขึ้น จึงใช้คำระยะช่องคำที่ 3 เป็นคำที่เหมาะสมในการตรวจหาคำนิเสธที่มีต่อคำแสดงความคิดเห็น

4.3 การสรุปบทวิจารณ์สินค้าจากผู้เชี่ยวชาญ

ในส่วนของการทดลองการสรุปบทวิจารณ์สินค้าจากผู้เชี่ยวชาญ เป็นการหาผลสรุปบทวิจารณ์สินค้าจากผู้เชี่ยวชาญ เพื่อที่จะนำมาใช้เปรียบเทียบกับผลสรุปที่ได้จากระบบเพื่อที่จะวัดความถูกต้องในการค้นหาคุณลักษณะและการระบุทิศทางของคุณลักษณะจากระบบ ซึ่งผลสรุปที่จะใช้ในการเปรียบเทียบ เป็นผลสรุปที่ได้จากผู้เชี่ยวชาญด้านการตลาด 3 ท่าน ซึ่งเป็นอาจารย์ที่รับผิดชอบงานสอนรายวิชาทางการตลาดของนิสิตระดับปริญญาบัณฑิตขึ้นไป โดยบทวิจารณ์

สินค้าที่นำมาให้ผู้เชี่ยวชาญสรุปจะเป็นชุดเดียวกันกับที่ใช้เป็นข้อมูลนำเข้าระบบ ซึ่งมีการรวบรวมบทวิจารณ์ของแต่ละสินค้าเป็นรูปเล่มให้ผู้เชี่ยวชาญอ่านและสรุปผลดังตัวอย่างในรูปที่ 2 ในเล่มบทวิจารณ์จะประกอบไปด้วยสินค้า 4 สินค้า และมีจำนวนบทวิจารณ์ในแต่ละสินค้าเป็นจำนวนเท่ากันกับที่ใช้ในระบบ โดยมีการแนบใบสรุปความคิดเห็นให้ผู้เชี่ยวชาญระบุคุณลักษณะของสินค้าที่ผู้บริโภคให้ความสนใจและระบุทิศทางของคุณลักษณะจากความคิดเห็นโดยรวมที่ผู้บริโภคกล่าวถึงคุณลักษณะในบทวิจารณ์ โดยแบ่งเป็นความคิดเห็นเชิงบวกและความคิดเห็นเชิงลบที่มีต่อคุณลักษณะนั้นๆ

หลังจากได้ผลสรุปบทวิจารณ์สินค้าจากผู้เชี่ยวชาญทั้ง 3 ท่านมาแล้ว ในกรณีที่ได้ผลสรุปไม่ตรงกัน จะนำผลการสรุปจากผู้เชี่ยวชาญทั้ง 3 มาเปรียบเทียบกัน เพื่อสรุปคุณลักษณะรวมและทิศทางของคุณลักษณะที่ได้ โดยมีเงื่อนไขว่าคุณลักษณะที่จะเป็นผลสรุปรวมจะต้องเป็นความเห็นที่ตรงกันของผู้เชี่ยวชาญอย่างน้อยจาก 2 ใน 3 ท่าน จากนั้นจะนำมาเปรียบเทียบทิศทางของคุณลักษณะที่สรุปได้ โดยเลือกผลสรุปทิศทางจากผู้เชี่ยวชาญอย่างน้อย 2 ท่าน เห็นตรงกันเช่นเดียวกัน ในกรณีที่มีคุณลักษณะที่ผู้เชี่ยวชาญเห็นตรงกัน 2 ใน 3 แต่ทิศทางของความคิดเห็นที่สรุปได้แตกต่างกัน ผู้วิจัยจะสรุปว่าไม่สามารถระบุทิศทางให้กับคุณลักษณะนั้นได้ ในการประเมินผล ผู้วิจัยจะนำเฉพาะผลสรุปที่ผ่านเงื่อนไขดังกล่าว มาใช้เปรียบเทียบกับผลสรุปจากระบบเท่านั้น โดยคุณลักษณะที่ไม่ผ่านเงื่อนไขดังกล่าวจะถูกตัดทิ้งไปและไม่นำมารวมในการประเมิน

4.4 การวัดผลการทดลอง

เมื่อได้ผลสรุปจากระบบและจากผู้เชี่ยวชาญแล้ว ผู้วิจัยนำผลสรุปทั้ง 2 มาเปรียบเทียบกัน เพื่อที่จะวัดความถูกต้องของระบบ ซึ่งมีการวัดผล 2 รูปแบบ ได้แก่ (1) การวัดความถูกต้องของระบบในการค้นหาคุณลักษณะของสินค้า จากการนำคุณลักษณะที่ได้จากระบบมาเปรียบเทียบกับคุณลักษณะที่ได้จากผู้เชี่ยวชาญ

(2) การวัดความถูกต้องของระบบในการระบุทิศทางให้กับคุณลักษณะของสินค้า โดยเปรียบเทียบจากคุณลักษณะที่ระบบระบุได้ตรงกับคุณลักษณะที่ได้จากผู้เชี่ยวชาญ ซึ่งในการวัดความถูกต้องของระบบสามารถวัดได้จากการคำนวณหาค่าระลึก (Recall) และค่าความแม่นยำ (Precision) โดยมีสูตรการคำนวณดังแสดงในสมการที่ 9 - 11

$$\text{ค่าระลึก (ในการค้นหาคุณลักษณะ)} = \frac{(\text{คุณลักษณะจากระบบที่สรุปตรงกับผู้เชี่ยวชาญ})}{\text{คุณลักษณะจากผู้เชี่ยวชาญทั้งหมด}} \quad (9)$$

$$\text{ค่าความแม่นยำ (ในการค้นหาคุณลักษณะ)} = \frac{(\text{คุณลักษณะจากระบบที่สรุปตรงกับผู้เชี่ยวชาญ})}{\text{คุณลักษณะจากระบบทั้งหมด}} \quad (10)$$

$$\text{ค่าความแม่นยำ (ในการระบุทิศทาง)} = \frac{(\text{จำนวนคุณลักษณะที่ระบบสามารถสรุปทิศทางตรงกับผู้เชี่ยวชาญ})}{\text{จำนวนคุณลักษณะที่ระบบสรุปตรงกับผู้เชี่ยวชาญ}} \quad (11)$$

5. ผลการทดลอง

ผลสรุปบทวิจารณ์สินค้าจากระบบและผู้เชี่ยวชาญ มีรายละเอียดดังแสดงในภาคผนวก โดยเมื่อเปรียบเทียบ ผลที่ได้แล้ว สามารถคำนวณความถูกต้องของระบบได้ดังแสดงในตารางที่ 4

ตารางที่ 4 ตารางสรุปผลการทดลอง

สินค้า	ความถูกต้องในการค้นหาคุณลักษณะ			ความถูกต้องในการระบุทิศทางของคุณลักษณะ
	ค่าระลึก	ค่าความแม่นยำก่อนกรอง	ค่าความแม่นยำหลังกรอง	ค่าความแม่นยำ
บริษัทออนทาแก้ม	0.56	0.28	0.63	0.80
โลชั่นทำความสะอาดหน้า	1.00	0.50	0.60	0.75
ครีมรองพื้น	0.67	0.28	0.45	0.25
ครีมรองพื้นผสมกันแดด	0.83	0.28	0.45	0.50
ค่าเฉลี่ย	0.77	0.33	0.53	0.58

อย่างไรก็ตาม เนื่องจากในบางครั้งคุณลักษณะที่ระบบสรุปได้เป็นคำที่ไม่มีความหมาย เช่น คำว่า “กก” ซึ่งความผิดพลาดดังกล่าวไม่ได้เกิดจากเทคนิคที่ใช้ในการพัฒนาระบบ แต่เกิดจากองค์ประกอบภายนอก เช่น ฐานข้อมูลคำศัพท์ที่ใช้ หรือลักษณะการเขียนที่ไม่ตรงตามไวยากรณ์ เช่น คำว่า “มากกกกก” ดังนั้นผู้วิจัยจึงสรุปค่าความแม่นยำแยกเป็นความแม่นยำก่อนและหลังการกรองความหมายของคำ

ระบบมีค่าความระลึกในการระบุคุณลักษณะของสินค้าเฉลี่ยเท่ากับ 0.77 ซึ่งอยู่ในระดับที่น่าพอใจ แต่มีค่าความแม่นยำเฉลี่ยไม่สูงมากนัก เนื่องจากคำแสดงคุณลักษณะที่จำแนกได้บางคำที่มีความหมายไม่เหมาะสมที่จะนำมาเป็นคำแสดงคุณลักษณะ หลังจากกรองคำที่มีความหมายไม่เหมาะสมออก ค่าความแม่นยำเฉลี่ยหลังกรองความหมายมีค่าเพิ่มขึ้นจาก 0.33 เป็น 0.53 ซึ่งผลต่างของค่าความแม่นยำก่อนและหลังการกรองความหมาย คาดว่าเกิดจาก (1) คำที่มาใช้เป็นคำแสดงคุณลักษณะบางคำเป็นคำที่มีหลายหน้าที่จากการพิจารณาคำที่มีความหมายไม่เหมาะสมพบว่า คำส่วนใหญ่เป็นคำกริยาที่มีหน้าที่คำมากกว่า 1 หน้าที่ คือนอกจากมีหน้าที่เป็นคำกริยาแล้วยังมีหน้าที่เป็นคำคุณศัพท์หรือคำกริยาวิเศษณ์ หรือทั้งสองอย่าง เช่นคำว่า “คือ” “มี” “สุด” ซึ่งคำเหล่านี้ขาดความเฉพาะเจาะจงในการแสดงคุณลักษณะของสินค้า (2) ฐานข้อมูลคำศัพท์ที่นำมาใช้ในการตัดคำไม่ได้พัฒนาขึ้นเพื่อใช้ในระบบสรุปบทวิจารณ์สินค้าโดยตรง เนื่องจากคำแสดงคุณลักษณะบางคำที่ควรเป็นคำเดียวกันแต่ถูกตัดเป็นสองคำ เช่นคำว่า “ติดทน” “เมื่อดีส” ดังนั้นหากสามารถปรับปรุงฐานข้อมูลของคำให้เหมาะสมกับงานในลักษณะนี้ จะทำให้ค่าความถูกต้องในการค้นหาคุณลักษณะของระบบสามารถเพิ่มขึ้นจาก 0.33 เป็น 0.53 ได้ และอาจจะส่งผลให้ค่าระลึกดีขึ้นด้วย

ส่วนความถูกต้องในการระบุทิศทางของคุณลักษณะมีค่าความแม่นยำเฉลี่ยอยู่ที่ 0.58 ปัจจัยที่คาดว่าส่งผลให้ค่าความแม่นยำไม่สูงมากอาจเกิดจาก 3 ส่วน ได้แก่ (1) การตรวจพบคำแสดงความคิดเห็นข้ามประโยค เช่น ประโยค “ตัวตลับสวยมากๆ ใช้สล็อตส้อมอยู่ตอนนี้” แบ่งออกได้เป็นประโยคที่ 1 คือ “ตัวตลับสวยมากๆ” มีคำว่า “ตลับ” เป็นคำแสดงคุณลักษณะและมีคำว่า “สวย” เป็นคำแสดงความคิดเห็น และประโยคที่ 2 คือ “ใช้สล็อตส้อมอยู่ตอนนี้” ซึ่งมีคำว่า “สล็อต” เป็นคำแสดงคุณลักษณะ แต่เมื่อตรวจไปทางซ้ายมือ 3 ช่องคำพบคำว่า “สวย” แต่คำว่า “สวย” ไม่ได้เป็นคำแสดงความคิดเห็นในประโยคที่สอง จึงทำให้เกิดความผิดพลาดในการระบุทิศทางของคุณลักษณะขึ้นได้ (2) การกำหนดทิศทางของคำในรายการแสดงความคิดเห็นบนคำที่มีความหมายคลุมเครือ เช่น คำว่า “ลื่น” ผู้บริโภคบางคนอาจชอบโลชั่นที่มีความลื่น ในขณะที่บางคนอาจไม่ชอบ (3) รายการคำแสดงความคิดเห็นต้นกำเนิดที่มีปริมาณน้อยอาจจะส่งผลให้ไม่พบคำในรายการต้นกำเนิดที่มีความสัมพันธ์กับคำที่ต้องการเพิ่มคำเข้ามาใหม่ ทำให้ไม่สามารถระบุทิศทางให้กับคำๆ นั้นได้ ซึ่งจะส่งผลกระทบต่อคำอื่นๆ ที่มีความสัมพันธ์กับคำนั้นต่อไป เมื่อคำแสดงความคิดเห็นในรายการมีน้อย อาจส่งผลให้ไม่สามารถระบุทิศทางกับคุณลักษณะบางอันได้ และไม่สามารถนำคุณลักษณะนั้นมาเปรียบเทียบเพื่อหาความถูกต้องของระบบในการระบุทิศทางของคุณลักษณะได้ ซึ่งหากสามารถลดความผิดพลาดหรือข้อจำกัดดังกล่าวลงได้ คาดว่าจะส่งผลให้ระบบมีความถูกต้องในการระบุทิศทางของคุณลักษณะมากยิ่งขึ้น

6. สรุปผลการวิจัย แนวทางนำไปประยุกต์ใช้ และข้อจำกัดของระบบ

ในการศึกษาและพัฒนาระบบสรุปบทวิจารณ์สินค้าแบบอัตโนมัติในการวิจัยครั้งนี้ พบว่า ค่าคงที่ต่างๆ ที่

เหมาะสมกับบทวิจารณ์ภาษาไทยที่ใช้ในการทดลอง มีดังนี้คือ ค่าเปอร์เซ็นต์ต่ำสุดเท่ากับ 0.1 ค่าปัจจัยการจำแนกเท่ากับ 1.5 โดยมีค่าระยะช่องคำในการตรวจหาคำแสดงความคิดเห็นที่มีต่อคุณลักษณะเท่ากับระยะช่องคำในการตรวจหาคำนิเสธที่มีต่อคำแสดงความคิดเห็นคือ 3 ช่องคำ

ในการค้นหาคุณลักษณะ คำระลึกที่ได้มีค่าสูงกว่าค่าความแม่นยำ ซึ่งในบริบทของการนำไปประยุกต์ใช้ ผู้วิจัยเห็นว่าคำระลึกมีความสำคัญมากกว่าค่าความแม่นยำ เนื่องจากคำระลึกเป็นค่าที่บ่งบอกถึงความถูกต้องของระบบเมื่อเทียบกับมนุษย์ ในขณะที่ค่าความแม่นยำเป็นค่าที่บอกความถูกต้องในการสรุปผลของระบบ เพราะฉะนั้นในระบบลักษณะนี้ซึ่งผู้ใช้ระบบเป็นมนุษย์ สามารถทำความเข้าใจกับความหมายของคำที่เป็นผลลัพธ์และเลือกใช้ผลลัพธ์ได้ คำระลึกที่ดีน่าจะมีความสำคัญต่อระบบมากกว่าค่าความแม่นยำ อย่างไรก็ตาม ความแม่นยำในการระบุทิศทางของคุณลักษณะก็มีความสำคัญไม่ยิ่งหย่อนไปกว่ากัน ซึ่งหากระบบสามารถระบุคุณลักษณะของสินค้าที่ผู้บริโภคให้ความสำคัญได้ถูกต้อง แต่ไม่สามารถระบุทิศทางความคิดเห็นของผู้บริโภคที่มีต่อคุณลักษณะนั้นๆ ได้อย่างแม่นยำ ก็จะไม่ส่งผลให้ไม่สามารถใช้ประโยชน์จากระบบได้อย่างเต็มที่ ดังนั้นจึงสรุปได้ว่าระบบที่พัฒนาขึ้นเป็นต้นแบบที่มีความแม่นยำในการระบุคุณลักษณะและทิศทางของความคิดเห็นในระดับพอใช้ และมีแนวทางในการนำไปประยุกต์ใช้ดังนี้

ด้วยวิธีการจำแนกคำที่นำมาใช้หาคุณลักษณะของคุณลักษณะของระบบที่พัฒนาขึ้น ในการนำระบบไปประยุกต์ใช้จำเป็นต้องมีบทวิจารณ์สินค้ามากกว่า 1 สินค้า และสามารถสรุปแนวทางในการประยุกต์ใช้ได้สองแนวทางคือ ในกรณีที่บทวิจารณ์สินค้าที่เป็นข้อมูลนำเข้าของระบบเป็นบทวิจารณ์สินค้าประเภทเดียวกัน เช่น ประเภทเครื่องสำอาง แต่ต่างกลุ่ม เช่น บทวิจารณ์สินค้าที่อยู่ใน

กลุ่มครีมรองพื้นและกลุ่มโลชั่น อาจจะมีคำบางคำที่เป็นคุณลักษณะร่วมระหว่างกลุ่มสินค้า จึงอาจทำให้เกิดการแย่งคำกัน ทำให้ไม่สามารถแยกคำนั้นออกมาเพื่อใช้เป็นคุณลักษณะของสินค้าทั้งในกลุ่มครีมรองพื้นและกลุ่มโลชั่น เพราะฉะนั้นสภาพแวดล้อมที่เหมาะสมที่จะนำวิธีการหาคุณลักษณะในงานวิจัยนี้ไปใช้ คือ การนำไปใช้กับสินค้าที่มีประเภทแตกต่างกัน เช่น สินค้าประเภทเครื่องสำอาง และสินค้าประเภทเครื่องใช้ไฟฟ้า เพราะจะมีคุณลักษณะแตกต่างกันโดยสิ้นเชิง ทำให้ไม่มีปัญหาเรื่องการแย่งคำเกิดขึ้น โดยสามารถนำคำที่จำแนกได้สำหรับสินค้าแต่ละประเภทไปใช้ในการวิเคราะห์บทวิจารณ์ของสินค้าประเภทนั้นๆ ได้โดยตรง แต่ในกรณีที่นำไปประยุกต์ใช้กับบทวิจารณ์สินค้าประเภทเดียวกัน และอยู่ในกลุ่มเดียวกัน เช่น บทวิจารณ์สินค้ากลุ่มบรืชออนเหมือนกันแต่ต่างยี่ห้อ สามารถนำคำแสดงคุณลักษณะที่ได้ของสินค้าแต่ละตัวมารวมกันเพื่อใช้เป็นส่วนหนึ่งของคำแสดงคุณลักษณะรวมในการวิเคราะห์สินค้าประเภทนั้นๆ ได้

งานวิจัยนี้ไม่รองรับการดำเนินการกับคำที่เขียนผิดไม่ว่าจะเป็นการสะกดผิดหรือการใช้คำผิด ในกรณีที่ผู้วิจารณ์สะกดคำผิด คำๆ นั้นจะถูกตัดออกในขั้นตอนการกำหนดหน้าที่คำ เนื่องจากเป็นคำที่ไม่พบในฐานข้อมูลคำศัพท์ ยกเว้นในกรณีที่ผู้วิจารณ์เขียนคำผิดแต่เป็นคำที่มีความหมาย เช่น เขียนคำว่า “เชี่ย” แทนที่จะเป็นคำว่า “เกลี่ย” ระบบจะตรวจสอบความหมายคำว่า “เชี่ย” ในฐานข้อมูลคำศัพท์ ถ้าพบว่า มีความหมายก็สามารถนำมาใช้เป็นคำแสดงความคิดเห็นได้ อย่างไรก็ตามในการหาคุณลักษณะของสินค้านั้น คำที่จะจำแนกออกมาเป็นคุณลักษณะได้ ต้องมีความถี่และการกระจายตัวของคำในบทวิจารณ์สินค้านั้นๆ สูง และโดดเด่นกว่าคำเดียวกันที่ปรากฏในบทวิจารณ์ของสินค้าอื่นๆ ดังนั้นคำที่เขียนผิดแต่เป็นคำที่มีความหมายจะมีโอกาสน้อยที่จะถูกแยกออกมาเป็นคำแสดงคุณลักษณะ

เอกสารอ้างอิง

- Aroonmanakun, W., Tansiri, K. and Nittayanuparp, P. 2009. **Thai National Corpus: A Progress Report**. Proceedings of the 7th Workshop on Asian Language Resources, ACL-IJCNLP 2009, pages 153-160, Suntec, Singapore, 6-7 August 2009.
- Charoenpornasawat, P. 1998. **Feature-based Thai Word Segmentation**. Master Thesis of Engineering. Department of Computer Engineering, Chulalongkorn University, Bangkok.
- Etzioni, M., Cafarella, D., Downey, S., Kok, A., Popescu, T., Shaked, S., Soderland, D. Weld, and A. Yates. 2005. **Unsupervised Named-Entity Extraction from the Web: An Experimental Study**. Artificial Intelligence, 165(1):91-134.
- Fellbaum, C. 1998. **WordNet : An Electronic Lexical Database**. MIT Press.
- Haruechaiyasak, C. and Kongyoung, S. 2009. **Thai Lexeme Analyser Based on the Conditional Random Fields: TLex**. Human Language Technology Laboratory (HLT), National Electronics and Computer Technology Center (NECTEC), Thailand.
- Hu, M. and Liu, B. 2004. **Mining and summarizing customer reviews**. In KDD : Proc. of the tenth ACM SIGKDD Conference on Knowledge discovery and data mining, pages 168-177. ACM.
- Meknavin, S., Charoenpornasawat, P. and Kijisirkul, B. 1997. **Feature-based Thai Word Segmentation**. In proceeding the Natural Language Processing Pacific Rim Symposium (NLPRS1997), page 41-46.
- Oelke, D., Hao, M., Rohrdantz, C., Keim, D., Dayal, U., Haug, L. and Janetzko, H. 2009. **Visual Opinion Analysis of Customer Feedback Data**. IEEE Symposium on Visual Analytics Science and Technology, Atlantic City, New Jersey, USA.
- Popescu, A. and Etzioni, O. 2005. **Extracting Product Features and Opinions from Reviews**. Department of Computer Science and Engineering, University of Washington Seattle, WA 98195-2350.
- Salton G., Buckley C.1988. **Term Weighting Approaches in Automatic Text Retrieval**. Information Processing and Management: An International Journal. 24:5.
- Werbler, C. and Harris, C. 2008. **Online Consumer Reviews Significantly Impact Consumer Purchasing Decisions**. Opinion Research Corporation and cjp communications.

กิตติกรรมประกาศ

นายตรีเทพ ชำรงลักษณ์ ไคร์ขอขอบพระคุณทุนสนับสนุนการค้นคว้าวิจัยจากมูลนิธิพระบรมราชานุสรณ์ พระบาทสมเด็จพระปกเกล้าเจ้าอยู่หัวและสมเด็จพระนางเจ้ารำไพพรรณี และคณะผู้วิจัยไคร์ขอขอบคุณ ผู้เชี่ยวชาญด้านการตลาด 3 ท่าน ได้แก่ อ.ดร.ณัฐพล อัสสระรัตน์ อ.ดร.พิมพ์ สุ่นสวัสดิ์ และ อ.ธนพร หงสุชน จากภาควิชาการตลาด คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ได้ให้ความอนุเคราะห์ในการสรุปคุณลักษณะและทิศทางความคิดเห็นที่ใช้ในการเปรียบเทียบความแม่นยำของระบบ

ภาคผนวก

ตารางที่ ก ผลสรุปทิศทางของคุณลักษณะของสินค้ากลุ่มบรัชอนทาแก้วที่ได้จากระบบ

สินค้า	คุณลักษณะ	บวก	ลบ	รวม	สรุปทิศทาง
บรัชอนทาแก้ว (ค่านาม)	สี	524	-43	481	+
	ติด	350	-17	333	+
	เม็ด	70	-5	65	+
	กก	46	-4	42	+
	แก้ว	19	-1	18	+
	ตลับ	15	-5	10	+
บรัชอนทาแก้ว (คำกริยา)	สี	355	-25	330	+
	กก	122	0	122	+
	สวย	119	-3	116	+
	ติด	115	-10	105	+
	ทน	106	-5	101	+
	ให้	93	-2	91	+
	เลือก	78	-1	77	+
	มี	51	-4	47	+
	อยู่	25	-4	21	+
	ปิด	65	-57	8	+
	สุด	10	-2	8	+
	คือ	4	-1	3	+

ตารางที่ ข ผลสรุปทิศทางของคุณลักษณะของสินค้ากลุ่มโลชั่นทำความสะอาดหน้าและเครื่องสำอางที่ได้จากระบบ

สินค้า	คุณลักษณะ	บวก	ลบ	รวม	สรุปทิศทาง
โลชั่นทำความสะอาด หน้าและเครื่องสำอาง (ค่านาม)	เครื่องสำอาง	20	-3	17	+
	สี	6	-4	2	+
	ราคา	7	-6	1	+
	กลิ่น	5	-7	-2	-
โลชั่นทำความสะอาด หน้าและเครื่องสำอาง (คำกริยา)	ล้าง	51	-2	48	+
	ออก	42	-1	41	+
	สะอาด	22	0	22	+
	ล้างหน้า	8	0	8	+
	อุดต้น	5	0	5	+
	รู้สึก	5	-1	4	+
	แพ้	4	-1	3	+
	แห้ง	3	-1	2	+

ตารางที่ ค ผลสรุปทิศทางของคุณลักษณะของสินค้ากลุ่มครีมรองพื้นที่ได้จากระบบ

สินค้า	คุณลักษณะ	บวก	ลบ	รวม	สรุปทิศทาง
ครีมรองพื้น (ค่านาม)	ผ่อง	36	0	36	+
	ขาว	7	-2	5	+
	แป้ง	5	0	5	+
	คน	5	-1	4	+
	รอย	3	-1	2	+
ครีมรองพื้น (คำกริยา)	ขึ้น	2	0	2	+
	ดู	1	0	1	+
	สว่าง	1	0	1	+

ตารางที่ ๖ ผลสรุปทิศทางของคุณลักษณะของสินค้ากลุ่มรองเท้าผสมกันแดดที่ได้จากระบบ

สินค้า	คุณลักษณะ	บวก	ลบ	รวม	สรุปทิศทาง
ครีมรองเท้าผสมกันแดด (คำนาม)	ผู้หญิง	29	0	29	+
	ชาย	5	-2	3	+
	แม่	3	0	3	+
	คน	6	-4	2	+
	รอย	1	-6	-5	-
ครีมรองเท้าผสมกันแดด (คำกริยา)	เกลียด	33	-2	31	+
	ดี	19	0	19	+
	ขึ้น	6	0	6	+
	คุม	7	-1	6	+
	เนียน	6	0	6	+
	ทา	8	-3	5	+
	คน	4	0	4	+
	ใส่	4	0	4	+
	ขาว	4	-1	3	+
	ช่วย	3	0	3	+
	ดู	3	-1	2	+
	สว่าง	1	-1	0	+
	หมอง	1	-2	-1	-