

การทำเหมืองความคิดเห็นภาษาไทย

CONSTRUCTING THAI OPINION MINING

มาสวีร์ มาศศิครโชติ

หัวหน้าสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร

คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยศรีปทุม

E-mail : masawee.ma@spu.ac.th

บทคัดย่อ

การเติบโตของเทคโนโลยีอินเทอร์เน็ตในปัจจุบันทำให้มนุษย์สามารถแลกเปลี่ยนความคิดเห็นผ่านสื่อสังคมออนไลน์ ฟอรัม หรือ เว็บบอร์ด ได้อย่างสะดวกรวดเร็วส่งผลให้จำนวนความคิดเห็นมีปริมาณเพิ่มขึ้นอย่างมหาศาล การทำเหมืองความคิดเห็นโดยอัตโนมัติสามารถนำมาใช้ประโยชน์ในด้านต่างๆ เช่น ช่วยในการตัดสินใจในการเลือกซื้อสินค้าให้กับผู้บริโภค และช่วยให้ผู้ประกอบการทราบถึงชื่อเสียงของสินค้าของตนเอง บทความนี้นำเสนอแนวคิดและหลักการการทำเหมืองความคิดเห็นที่มุ่งเน้นประเด็นการวิเคราะห์ความคิดเห็นแบบปริยายสำหรับข้อความภาษาไทยเพื่อหาข้อสรุปว่าความคิดเห็นเหล่านั้นแสดงความรู้สึกในเชิงบวกหรือเชิงลบ การวิเคราะห์และสังเคราะห์งานวิจัยที่เกี่ยวข้องแสดงให้เห็นการวิเคราะห์ความคิดเห็นในระดับคุณลักษณะมีความละเอียดมากกว่าระดับอื่น และการวิเคราะห์ความคิดเห็นภาษาไทยด้านการวิเคราะห์ความคิดเห็นที่ชัดเจนยังมีข้อจำกัด กล่าวคือ ความคิดเห็นที่นำมาวิเคราะห์ต้องมีคำระบุข้อความที่ชัดเจน จึงนำมาสู่การเสนอแนะแนวทางการวิเคราะห์ความคิดเห็นแบบปริยายสำหรับข้อความภาษาไทย โดยการสกัดคำหรือวลีในระดับคุณลักษณะของวัตถุร่วมกับการพิจารณาบริบทในบริเวณใกล้เคียง

คำสำคัญ : การทำเหมืองความคิดเห็น การวิเคราะห์ความคิดเห็น ความคิดเห็นแบบปริยาย

ABSTRACT

Nowadays the growth of Internet technology makes people share opinions in social media, forum and web boards easily. This causes the huge volume of opinionated text. Opinion mining aims for analyzing opinion sentences for various purposes such as to help consumers in decision makings and to allow entrepreneurs to know the reputations of their products. This article reviews opinion mining concepts and principle that focuses on sentiment analysis in implicit opinions for Thai language to summarize the polarities of opinions that express positive or negative sides. The analysis and synthesis of related research has shown that feature-based sentiment analysis provides finer grains than the other levels, and the sentiment analysis in explicit opinion for Thai language still has a limitation in that polar words have to be expressed explicitly in sentences. This leads to the suggestion of sentiment analysis in implicit opinions for Thai language by using words or phrases extraction in feature level together with the determination of surrounding context.

KEYWORDS: Opinion Mining, Sentiment Analysis, Implicit Opinion?

บทนำ

ความก้าวหน้าของเทคโนโลยีอินเทอร์เน็ตนำมาซึ่งการเปลี่ยนแปลงวิธีการแสดงมุมมองและความรู้สึกของมนุษย์ โดยช่วยให้สามารถสร้างเนื้อหา แบ่งปัน แลกเปลี่ยนความรู้สึก ความคิดเห็น และประสบการณ์ ได้อย่างสะดวกรวดเร็วผ่านสื่อสังคมออนไลน์ ฟอรัม หรือเว็บบอร์ด เช่น การแสดงความคิดเห็นของผู้บริโภคเกี่ยวกับสินค้าและการให้บริการและการแสดงความคิดเห็นเกี่ยวกับเหตุการณ์ทางการเมืองหรือข่าวต่างๆ ก่อให้เกิดแหล่งรวบรวมความคิดเห็นขนาดใหญ่ ช่วยให้ทั้งผู้บริโภคและผู้ประกอบการสามารถนำข้อคิดเห็นเหล่านี้ไปใช้ประโยชน์ในด้านต่างๆ อาทิ ช่วยให้ผู้บริโภคทราบข้อมูลจากผู้ใช้งานจริงเพื่อสามารถตัดสินใจเลือกซื้อสินค้าหรือบริการที่ต้องการ และช่วยให้ผู้ประกอบการเข้าถึงชื่อเสียงของสินค้าหรือบริการทั้งของตนเองและคู่แข่งเพื่อนำไปวางแผนพัฒนาหรือปรับปรุงคุณภาพสินค้าและบริการให้ตรงกับความต้องการของผู้บริโภค

แต่ในภาวะการณ์ปัจจุบันการแสดงความรู้สึกหรือแลกเปลี่ยนประสบการณ์เกี่ยวกับสินค้าหรือบริการโดยมักไม่จำกัดเพศ อายุ หรือสถานภาพทางสังคม ทำให้เนื้อหาที่ปรากฏในแต่ละข้อ ความแตกต่างกันตามมุมมองหรือความรู้สึกของแต่ละคน ประกอบกับการเปิดกว้างในการแสดงความคิดเห็นจึงส่งผลให้จำนวนความคิดเห็นมีปริมาณเพิ่มขึ้นอย่างรวดเร็วและมีปริมาณมหาศาล เช่นความคิดเห็นเกี่ยวกับโทรศัพท์เคลื่อนที่จึงเป็นการยากที่ผู้บริโภคจะสามารถเข้าถึงข้อมูลทั้งหมดและนำข้อมูลเหล่านั้นมาใช้ประกอบการตัดสินใจได้ ดังนั้นสิ่งที่ผู้บริโภคหรือผู้ประกอบการต้องการคือวิธีการสกัด ความคิดเห็นจากข้อมูลปริมาณมากโดยอัตโนมัติเพื่อนำมาสรุปว่าความคิดเห็นเหล่านั้นแสดงความรู้สึกไปในเชิงบวก หรือเชิงลบ ซึ่งจะสามารถนำมาช่วยในการตัดสินใจ โดยทั้งฝ่ายผู้บริโภคและฝ่ายผู้ประกอบการจะได้รับทราบข้อมูลที่สะท้อนให้เห็นถึงมุมมองและความรู้สึกที่แท้จริงที่มีต่อสินค้าหรือบริการนั้นๆ และสามารถนำข้อมูลเหล่านี้ไปใช้ประโยชน์ในด้านต่างๆ ยกตัวอย่างเช่น ความคิดเห็นที่แสดงความรู้สึกในเชิงบวกสามารถช่วยดึงดูดให้ผู้บริโภคตัดสินใจซื้อสินค้าหรือบริการ ในทางกลับกันความคิดเห็นที่แสดงความรู้สึกในเชิงลบ สามารถช่วยให้ผู้ประกอบการหันกลับมาปรับปรุงคุณภาพของสินค้าหรือบริการให้ดียิ่งขึ้น ซึ่งสามารถช่วยประหยัดทั้งในเรื่องของเวลา และงบประมาณมากกว่าการจ้างบริษัทที่ปรึกษาหรือวิธีการสำรวจความคิดเห็นแบบเดิม

เทคนิคในการวิเคราะห์สรุปความคิดเห็นโดยอัตโนมัติ ที่ได้รับความนิยมอย่างต่อเนื่องในช่วงหลายปีที่ผ่านมา คือ การทำเหมืองความคิดเห็น (opinion mining) หรือการวิเคราะห์ความคิดเห็น (sentiment analysis) ซึ่งเป็นแนวคิดหนึ่งในศาสตร์แห่งการประมวลผลภาษาธรรมชาติ (natural language processing: NLP) ที่สามารถใช้วิเคราะห์ความรู้สึก ทั้งในระดับเอกสาร ระดับประโยค และในระดับคุณลักษณะของวัตถุ รวมถึงการใช้วิเคราะห์ความคิดเห็นของผู้บริโภคที่มีความรู้สึกต่อสินค้าและบริการต่างๆ

บทความนี้นำเสนอแนวคิดสำคัญ และความก้าวหน้าของการประมวลผลภาษาธรรมชาติโดยมุ่งเน้นประเด็นการวิเคราะห์การทำเหมืองความคิดเห็นภาษาไทยแบบปริยายที่แสดงให้เห็นถึงการวิเคราะห์จุดเด่นและข้อจำกัดของงานวิจัยที่เกี่ยวข้องเพื่อสังเคราะห์ภาพรวม ในแต่ละแขนงของการทำเหมืองความคิดเห็น และยังรวมถึงการสังเคราะห์แนวทางที่เป็นไปได้ในการแก้ปัญหาการทำเหมืองความคิดเห็นภาษาไทยที่มีอยู่ด้วยการสกัดคำหรือวลีในระดับคุณลักษณะของวัตถุร่วมกับการพิจารณาบริบทในบริเวณใกล้เคียงกับคุณลักษณะนั้นในประโยคเพื่อนำไปสู่การระบุข้อความความคิดเห็น

การประมวลผลภาษาธรรมชาติ

การประมวลผลภาษาธรรมชาติเป็นวิธีการที่จะช่วยทำให้คอมพิวเตอร์เข้าใจภาษาธรรมชาติของมนุษย์ที่ไม่มีโครงสร้างที่แน่นอน โดยการแปลให้อยู่ในรูปแบบที่เป็นโครงสร้างการประมวลผลภาษาธรรมชาติสามารถแบ่งตามระดับการวิเคราะห์ได้ดังนี้ (บุญเสริม กิจศิริกุล, 2546:92)

1. การวิเคราะห์เชิงองค์ประกอบ (morphological analysis) คือการวิเคราะห์ในระดับของคำ เช่น คำว่า การประปะปาก็จะแยกเป็น คำว่า การ กับ คำว่า ประปะปาก็จะแยกเป็น คำว่า การ กับ คำว่า ประปะป

2. การวิเคราะห์ระดับโครงสร้างทางภาษาหรือวากยสัมพันธ์ (syntactic analysis) คือ การวิเคราะห์คำตามหน้าที่ของคำ (part-of-speech) จุดมุ่งหมายเพื่อต้องการดูว่าประโยคซึ่งประกอบด้วย คำหลายๆ คำเรียงต่อกันนั้นมีโครงสร้างทางภาษาเป็นอย่างไร คำใดเป็นประธาน กริยา กรรม หรือส่วนใดเป็นวลี เป็นต้น

3. การวิเคราะห์ระดับความหมาย (semantic analysis) คือการวิเคราะห์ความหมายของคำในประโยค

4. บูรณาการทางวาทศิลป์ (discourse integration) เป็นการวิเคราะห์ความหมายของประโยคโดยพิจารณาจากประโยคข้างเคียง

5. การวิเคราะห์ทางปฏิบัติ (pragmatic analysis) คือการแปลความหมายของประโยคใหม่อีกครั้งว่าผู้พูดตั้งใจจะสื่อความหมายอะไร

การวิเคราะห์ภาษาธรรมชาติในภาษาไทยส่วนใหญ่เป็นการวิเคราะห์ทางองค์ประกอบและการวิเคราะห์ระดับโครงสร้างทางภาษาหรือวาทศิลป์เท่านั้น เนื่องจากเครือข่ายคำไทย (Thai Wordnet) (Kongthon et al., 2010) ไม่ครอบคลุมคำศัพท์เฉพาะบางคำ

หัวข้อถัดไปเป็นการนำเสนอแนวคิด และวิธีการทำเหมืองความคิดเห็นที่ปรากฏอยู่ในปัจจุบัน

การทำเหมืองความคิดเห็น (Opinion Mining) หรือ การวิเคราะห์ความคิดเห็น (Sentiment Analysis)

ความคิดเห็น (opinion) หมายถึง การแสดงทัศนคติที่ผู้เขียนมีต่อสิ่งๆ หนึ่งหรือหัวข้อๆ หนึ่งโดยแบ่งออกเป็น (วรัญญา วรณศรี, 2553)

1. ความคิดเห็นแบบตรง (direct opinion) คือความคิดเห็นที่สามารถบอกทัศนคติของผู้เขียนหรือผู้พูดได้อย่างชัดเจน เช่น "โรงแรมนี้สะอาดมาก"

2. ความคิดเห็นเชิงเปรียบเทียบ (comparative opinion) คือ ความคิดเห็นที่เกิดจากการนำสิ่งที่ต้องการแสดงความคิดเห็นไปเปรียบเทียบกับอีกสิ่งหนึ่ง เช่น "โรงแรม A สะอาดกว่า โรงแรม B"

การทำเหมืองความคิดเห็นมีวัตถุประสงค์ในการสกัดข้อความที่เป็นความคิดเห็นที่มีต่อสิ่งใดสิ่งหนึ่งออกมาจากข้อเท็จจริงเพื่อนำไปสกัดคุณลักษณะและส่วนประกอบของวัตถุที่ถูกแสดงความคิดเห็นออกมา (Hu and Lui, 2004) เพื่อพิจารณาว่าความคิดเห็นนั้นแสดงความรู้สึกไปในเชิงบวก เชิงลบ หรือเป็นกลาง และสรุปข้อมูลที่ต้องการ สามารถแบ่งตามการวิเคราะห์ความคิดเห็นออกเป็น 3 ระดับ ดังนี้คือ (Liu, 2012)

1. การวิเคราะห์ความคิดเห็นในระดับเอกสาร (document-based sentiment analysis) เป็นการวิเคราะห์ข้อความแสดงความคิดเห็นทั้งหมดจากเอกสารและนำมาสรุป เพื่อทำการระบุข้อความความคิดเห็น (polar word) ว่าเป็นความคิดเห็นเชิงบวก เชิงลบ หรือเป็นกลาง ตัวอย่างการศึกษากการวิเคราะห์ความคิดเห็นในระดับเอกสาร คืองานวิจัยของ Turney (Turney, 2002) และงานวิจัยของ Pang et al. (Pang et al., 2002)

2. การวิเคราะห์ความคิดเห็นในระดับประโยค (sentence-based sentiment analysis) ในกรณีที่ภายในเอกสารประกอบด้วยประโยคที่มีความคิดเห็นที่แตกต่างกัน การวิเคราะห์ความคิดเห็นในระดับเอกสารจะไม่สามารถทำการสรุปข้อความความคิดเห็นได้ จึงจำเป็นต้องทำการวิเคราะห์ความคิดเห็นในระดับ ประโยคโดยแยกประโยคที่เป็นข้อความแสดงความคิดเห็นมาทำการระบุข้อความความคิดเห็นว่าเป็นความคิดเห็นเชิงบวก เชิงลบ หรือเป็นกลาง ตัวอย่างการศึกษากการวิเคราะห์ความคิดเห็นในระดับประโยค คืองานวิจัยของ Soo-Min Kim และ Eduard Hovy (Kim and Hovy, 2004)

3. การวิเคราะห์ความคิดเห็นในระดับคุณลักษณะของวัตถุ (feature-based sentiment analysis) ในกรณีที่ภายในประโยคเดียวกันประกอบด้วยความคิดเห็นที่มีต่อสิ่งใดสิ่งหนึ่งแตกต่างกัน การวิเคราะห์ความคิดเห็นในระดับประโยคจะไม่สามารถนำมาใช้ได้เนื่องจากไม่มีความละเอียดมากพอ จึงต้องทำการวิเคราะห์ความคิดเห็นในระดับคุณลักษณะของวัตถุ เพื่อทำการแยกคุณลักษณะที่สนใจของวัตถุ (feature) ในประโยคมาทำการระบุข้อความความคิดเห็นว่าเป็นความคิดเห็นเชิงบวก เชิงลบหรือเป็นกลาง ตัวอย่างการศึกษากการวิเคราะห์ความคิดเห็นในระดับคุณลักษณะของวัตถุ คือ งานวิจัยของ Hu and Lui (2004) และงานวิจัยของ Kongthon et al. (Kongthon et al., 2011)

ในปัจจุบันการวิเคราะห์ความคิดเห็นในบทวิจารณ์ทั้งภาษาไทยและภาษาต่างประเทศมุ่งเน้นที่การวิเคราะห์ความคิดเห็นในระดับคุณลักษณะของวัตถุ เนื่องจากเป็นการวิเคราะห์ข้อความความคิดเห็นที่ละเอียดกว่าระดับอื่น ดังอธิบายโดยละเอียดในหัวข้อถัดไป

การวิเคราะห์ความคิดเห็นในระดับคุณลักษณะของวัตถุ (Feature-based Sentiment Analysis)

การวิเคราะห์ความคิดเห็นในระดับคุณลักษณะของวัตถุสามารถแบ่งการทำงานเป็น 2 ขั้นตอน ดังนี้

1. การสกัดคุณลักษณะ (feature extraction) คือ การตรวจหาและนำคุณลักษณะ (feature) ของวัตถุที่มีการกล่าวถึงในเอกสารออกมา โดยนิยมใช้คำว่าคุณลักษณะในกรณีที่เป็นโดเมนของสินค้า แต่ในโดเมนอื่น วัตถุซึ่งเป็นเหตุการณ์หรือหัวข้อต่าง ๆ เช่น เหตุการณ์ทางการเมือง จะนิยมใช้คำว่ามุมมอง (aspect) แทนคำว่าคุณลักษณะ

2. การกำหนดทิศทางความคิดเห็น (opinion orientation) หรือการระบุขั้วความคิดเห็น (polarity of opinion) คือ การกำหนดทิศทางความคิดเห็นหรือการระบุขั้วความคิดเห็นที่มีต่อคุณลักษณะของวัตถุที่ได้จากขั้นตอนแรกว่าเป็นความคิดเห็นเชิงบวก เชิงลบหรือเป็นกลาง

ในส่วนของรูปแบบการวิเคราะห์ความคิดเห็นในระดับคุณลักษณะของวัตถุ สามารถแบ่งได้เป็น 2 รูปแบบ ดังนี้

1. การวิเคราะห์ความคิดเห็นที่ชัดเจน (explicit opinion)

การวิเคราะห์ความคิดเห็นที่ชัดเจนเป็นการวิเคราะห์ความคิดเห็นในระดับคุณลักษณะของวัตถุที่มักใช้คำนามหรือคำนามวลีที่อยู่ในประโยคมาเป็นคุณลักษณะของวัตถุหรือหัวข้อที่ถูกแสดงความคิดเห็น และใช้คำคุณศัพท์ (adjective) ที่อยู่ในประโยคเดียวกันเป็นคำแสดงความคิดเห็นซึ่งคำเหล่านี้จัดอยู่ในประเภทของคำที่สะท้อนถึงความรู้สึกได้อย่างชัดเจนยกตัวอย่างเช่น "อาหารร้านนี้อร่อยมาก" คำว่า "อาหาร" คือคุณลักษณะของร้าน "อร่อย" คือ คำแสดงความคิดเห็นที่สามารถระบุขั้วความคิดเห็นว่าเป็นไปในเชิงบวก หรือ "การบริการของโรงแรมนี้แย่มาก" คำว่า "การบริการ" คือคุณลักษณะของโรงแรม "แย่" คือ คำแสดงความคิดเห็นที่สามารถระบุขั้วความคิดเห็นว่าเป็นไปในเชิงลบ เป็นต้น แต่เนื่องจากกลุ่มคำแสดงความคิดเห็นเป็นกลุ่มคำที่ขึ้นอยู่กับโดเมน (domain-dependent) คำแสดงความคิดเห็นที่อยู่ในโดเมนหนึ่งอาจให้ขั้วความคิดเห็นที่แตกต่างกันเมื่ออยู่ในอีกโดเมนหนึ่ง เช่น คำว่า "ยาวนาน" สามารถระบุขั้วความคิดเห็นว่าเป็นไปในเชิงบวกสำหรับคุณลักษณะ "แบตเตอรี่" ในโดเมนโทรศัพท์มือถือ (นั่นคือ อายุการใช้งานของแบตเตอรี่ยาวนาน)

แต่สามารถระบุขั้วความคิดเห็นว่าเป็นไปในเชิงลบสำหรับคุณลักษณะ "การบริการ" ในโดเมนร้านอาหาร (นั่นคือใช้เวลายาวนานในการรอรับบริการ) ดังนั้น ในโดเมนที่แตกต่างกันจำเป็นต้องทำเหมืองความคิดเห็นที่ต่างกัน ตัวอย่างงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความคิดเห็นที่ชัดเจนมีดังนี้

Hu and Lui (2004) นำเสนอวิธีการสรุปผลความคิดเห็นในระดับคุณลักษณะ (feature-based opinion summarization) จากบทวิจารณ์สินค้าในเว็บขายสินค้า โดยแบ่งการทำงานของระบบออกเป็น 2 ขั้นตอนคือ ขั้นตอนของการสกัดคุณลักษณะที่สนใจของวัตถุหรือหัวข้อที่ถูกแสดงความคิดเห็น โดยใช้วิธีการทำเหมืองข้อมูลแบบหาความสัมพันธ์ (association mining) ในการหาคำแสดงคุณลักษณะของสินค้า และขั้นตอนที่สองคือการระบุขั้วความคิดเห็นว่าเป็นขั้วบวก ขั้วลบ หรือเป็นกลาง โดยใช้เทคนิคการแบ่งประเภทความคิดเห็นในบทวิจารณ์ด้วยวิธีการใช้เครือข่ายคำซึ่งอาศัยวิธีการอ้างอิงจากพจนานุกรม (dictionary-based approach) มาช่วยในการหาความสัมพันธ์ของคำ เพื่อแบ่งแยกคำแสดงความคิดเห็นว่าเป็นเชิงบวก หรือเชิงลบ แต่วิธีการอ้างอิงจากพจนานุกรมไม่สามารถระบุคำแสดงความคิดเห็นที่เป็นกลุ่มคำที่ขึ้นอยู่กับโดเมนได้ เนื่องจากคำศัพท์ในพจนานุกรมเป็นกลุ่มคำที่ไม่ขึ้นอยู่กับโดเมน นอกจากนี้แนวคิดดังกล่าวไม่ได้พิจารณาถึงตำแหน่งของคำในประโยค ต่อมา Qui et al. (2009) ได้นำเสนอวิธีการที่เรียกว่า Double Propagation โดยอาศัยวิธีการอ้างอิงจากคลังคำศัพท์ (corpus-based approaches) และพิจารณาถึงโครงสร้างทางไวยากรณ์ที่เกี่ยวข้องกับการวางตำแหน่งของคำในประโยคมากกว่าวิธีการของ Hu และ Lui โดยใช้ความสัมพันธ์แบบภาวะพึ่งพาในการสกัดคำแสดงความคิดเห็นและคุณลักษณะของสินค้า แต่วิธีการนี้ไม่เหมาะกับโดเมนขนาดใหญ่และให้ค่าความแม่นยำต่ำ Zhang et al. (2010) จึงเสนอวิธีการที่ช่วยแก้ปัญหาที่เกิดขึ้นในงานวิจัยของ Qui et al. โดยการใช้รูปแบบที่เรียกว่า part-whole และ "no" เพื่อช่วยเพิ่มการระลึก (recall) และใช้การจัดอันดับคุณสมบัติเพื่อช่วยเพิ่มค่าความแม่นยำ (precision) แต่วิธีการดังกล่าวเป็นวิธีการหาคุณลักษณะแบบทีละประโยคต่อประโยคไม่สามารถใช้กับบทวิจารณ์ที่เป็นภาษาไทยได้ เนื่องจากภาษาไทยไม่สามารถหาจุดสิ้นสุดของประโยคที่ชัดเจนได้เหมือนภาษาอังกฤษ

สำหรับการวิเคราะห์ความคิดเห็นภาษาไทยจะมีกระบวนการที่ยุ่ยากและซับซ้อนกว่าภาษาอังกฤษ เนื่องจากภาษาไทยเป็นภาษาที่ไม่มีขอบเขตของประโยคไม่เหมือนในภาษาอังกฤษ การแบ่งประโยคหรือแบ่งคำภาษาไทยไม่มีรูปแบบที่ตายตัว ไม่มีโครงสร้างที่ชัดเจนทำให้เกิดความกำกวม โดย ตรีเทพ อารังลักษณ์ และ จันทรเจ้า มงคลนาวัน (2554) พัฒนาระบบสรุปบทวิจารณ์สินค้าภาษาไทยโดยผู้บริโภคแบบอัตโนมัติ จากเว็บไซต์จีบัน (<http://www.jebun.com>) ซึ่งเป็นเว็บไซต์สำหรับแลกเปลี่ยนความคิดเห็นเกี่ยวกับเครื่องสำอางค์ โดยใช้วิธีการของ Oelke et al. (Oelke et al., 2009) ในการพิจารณาหาความเหมาะสมของคำแสดงคุณลักษณะสำหรับใช้สรุปบทวิจารณ์สินค้าที่มีข้อความภาษาไทยและในส่วนของการทำงานคำแสดงความคิดเห็นใช้วิธีการแบ่งประเภทความคิดเห็นด้วยวิธีการใช้เครือข่ายคำ จากนั้นสรุปคุณลักษณะของสินค้าที่ผู้บริโภคให้ความสำคัญ ทิศทางของคำแสดงความคิดเห็นที่มีต่อคุณลักษณะนั้นๆ ของสินค้า และเปรียบเทียบผลสรุปบทวิจารณ์สินค้าที่ได้จากระบบ กับผลสรุปบทวิจารณ์จากผู้เชี่ยวชาญทางด้านการตลาด ผลที่ได้คือระบบมีความแม่นยำในการระบุคุณลักษณะและทิศทางของความคิดเห็นในระดับพอใช้ ซึ่งจากการใช้เครือข่ายคำทำให้ไม่สามารถระบุคำแสดงความคิดเห็นที่เป็นกลุ่มคำที่ขึ้นอยู่กับทอเมนได้ Haruechaiyasak et al. (2010) จึงนำเสนอกรอบการทำงานในการสร้างคลังคำศัพท์ในการทำเหมืองความคิดเห็นภาษาไทย สำหรับโรงแรมโดยการวิเคราะห์ระดับโครงสร้างทางภาษา ซึ่งประกอบด้วยกลุ่มคำที่ขึ้นอยู่กับทอเมนและกลุ่มคำที่ไม่ขึ้นอยู่กับทอเมน ต่อมา Kongthon et al. (2011) ใช้วิธีการอ้างอิงจากคลังคำศัพท์ดังกล่าวในการพัฒนาระบบวิเคราะห์ข้อความ แสดงความคิดเห็นที่เป็นภาษาไทยสำหรับโรงแรม โดยอาศัยหลักการวิเคราะห์และสรุปผลจากความรู้สึกในระดับคุณลักษณะ (feature-based opinion analysis and summarization) ซึ่งการทำงานของระบบแบ่งเป็น 2 ขั้นตอน คือ ขั้นตอนการสกัดคุณลักษณะที่สนใจของวัตถุหรือหัวข้อที่ถูกแสดงความคิดเห็น และขั้นตอนการระบุข้อความความคิดเห็นว่าเป็นขั้วบวกหรือขั้วลบ และนำเสนอความคิดเห็นอยู่ในรูปแบบกราฟิกที่เข้าใจง่าย แต่แนวคิดดังกล่าวยังจำเป็นต้องใช้คำหรือวลีเป็นกลุ่มคำแสดงความคิดเห็นที่สะท้อนความรู้สึกได้อย่างชัดเจน เช่น ดี แย่

2. การวิเคราะห์ความคิดเห็นแบบปริยาย (implicit opinion)

การวิเคราะห์ความคิดเห็นแบบปริยายเป็นการวิเคราะห์ความคิดเห็นในระดับคุณลักษณะของวัตถุที่คำหรือวลีบางประเภทที่ไม่ได้จัดอยู่ในประเภทของคำที่สะท้อนถึงความรู้สึก (opinion word) แต่เมื่อใดก็ตามที่คำเหล่านั้นปรากฏในประโยคคำเหล่านั้นก็จะสามารถบอกโดยปริยายได้ว่าความคิดเห็นนั้นแสดงความรู้สึกไปในเชิงบวกหรือลบโดยอาจไม่จำเป็นต้องมีคำคุณศัพท์ในการระบุความคิดเห็น เช่น กลุ่มของคำที่หมายถึงคุณลักษณะของสินค้า ซึ่งกลุ่มคำเหล่านี้มักปรากฏในบทวิจารณ์สินค้าหลากหลายประเภท อาทิ สินค้าประเภทอุปกรณ์เคลื่อนที่ ยกตัวอย่างเช่น คำว่า "แบตเตอรี่" คือ คำแสดงคุณลักษณะของโทรศัพท์มือถือ เมื่อไปปรากฏในข้อความแสดงความคิดเห็น "โทรศัพท์มือถือเครื่องนี้กินแบตเตอรี่" ซึ่งไม่มีกลุ่มคำแสดงความคิดเห็นอยู่ในประโยค (คำว่า "กิน" เป็นคำกริยาซึ่งไม่ใช่คำแสดงความคิดเห็น) แต่สามารถบอกโดยปริยายว่าความคิดเห็นที่มีต่อโทรศัพท์นั้นเป็นไปในเชิงลบ หรือ "ในช่วงเดือนที่ผ่านมา พบว่ามีแอ่ง (valley) บนฟูกที่นอน" ซึ่งไม่มีกลุ่มคำแสดงความคิดเห็นอยู่ในประโยคซึ่งในที่นี้คำว่า "แอ่ง" หมายถึง คุณภาพของฟูกที่นอนและบอกโดยปริยายว่าความคิดเห็นที่มีต่อฟูกที่นอนเป็นไปในเชิงลบ ตัวอย่างงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความคิดเห็นแบบปริยาย มีดังนี้

Riloff et al. (2003) เสนอวิธีการสกัดคำนามที่มาจากคำกริยา (subjective noun) เช่น sleeper เป็นคำนามที่มาจากคำกริยา sleep แต่วิธีการดังกล่าวไม่สามารถใช้ในกรณีของการสกัดคำนามที่บอกความหมายโดยปริยายว่าเป็นความคิดเห็นได้ เนื่องจากคำนามที่มาจากคำกริยาอาจไม่ใช่คำที่บอกความหมายโดยปริยายว่าเป็นความคิดเห็น ต่อมา Popescu and Etzioni (2005) ได้คิดค้นวิธีโอพีไอเอ็นอี (OPINE) ซึ่งเป็นการดึงคุณลักษณะของสินค้าและระบุข้อความความคิดเห็นจากบทวิจารณ์สินค้าด้วยการคำนวณค่า PMI (Pointwise Mutual Information) ซึ่งในครั้งแรกใช้รูปแบบ part-whole ในการทำเหมืองคุณลักษณะ จุดเด่นของงานวิจัยนี้คือสามารถระบุคุณลักษณะโดยปริยาย (implicit feature) ได้ เช่น คำว่า "good scans" ซึ่งมีความหมายโดยปริยายถึง คุณภาพในการกราดภาพดี Zhang and Liu (2011a) ได้นำเสนอวิธีการในการระบุคุณลักษณะของสินค้าที่สามารถนำมาใช้สะท้อนถึงความรู้สึกที่มีต่อสินค้า และใช้กลุ่มของคำศัพท์แสดงความคิดเห็น (opinion lexicon) ของ Ding et al. (2008)

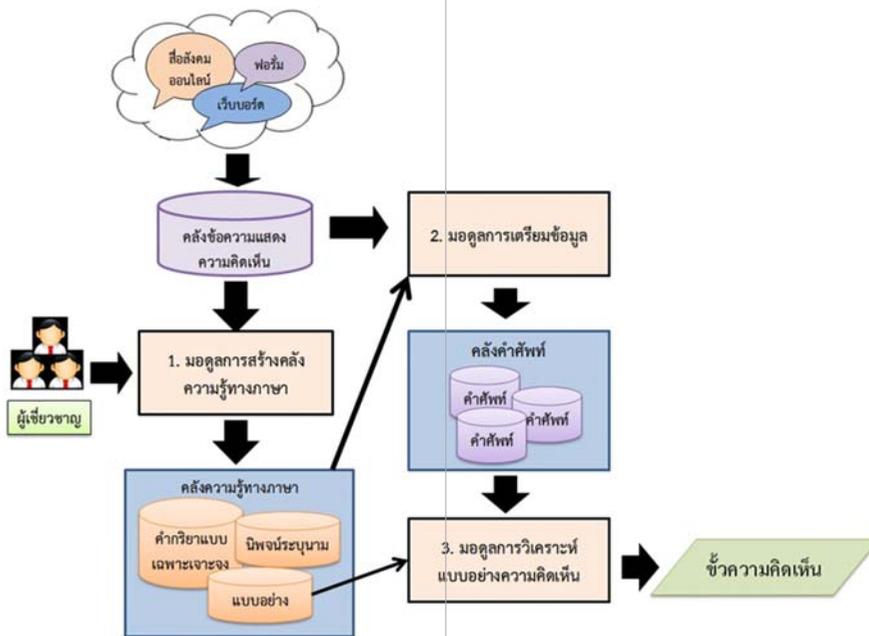
ในการระบุข้อความความคิดเห็นว่าเป็นเชิงบวกหรือลบ และพบว่าถ้าไม่มีการระบุคำศัพท์แสดงความคิดเห็นวิธีการดังกล่าวจะให้ค่าระลอกต่ำ ต่อมา Zhang and Liu (2011b) นำเสนอวิธีการที่ไม่จำเป็นต้องมีคำศัพท์แสดงความคิดเห็นในประโยคโดยการสกัดคำ หรือวลีที่อยู่ในรูปแบบของทรัพยากร (resource) เช่น น้ำ ไฟ แก๊ส ที่สามารถนำมาใช้สะท้อนถึงความรู้สึกที่มีต่อสินค้าในเอกสารความคิดเห็นที่เป็นภาษาอังกฤษ ซึ่งคำเหล่านี้ไม่ได้จัดอยู่ในประเภทคำที่สะท้อนถึงความรู้สึกแต่เมื่อไรก็ตามที่คำเหล่านั้นไปปรากฏในประโยค คำเหล่านี้จะสามารถบอกโดยปริยายได้ว่าความคิดเห็นนั้นแสดงความรู้สึกไปในเชิงบวกหรือลบ เช่น " This washer uses a lot of electricity." ในที่นี้คำว่า "electricity" เป็นทรัพยากรที่สะท้อนถึงความรู้สึกในเชิงลบที่มีต่อ "washer" เป็นต้น โดยได้นำเสนอวิธีการที่เรียกว่า MRE (Mutual Reinforcement based on Expected values) สำหรับการสะท้อนความสัมพันธ์ระหว่างคำหรือวลีในรูปแบบทรัพยากรกับคำกริยาที่เกี่ยวข้องในการสกัดคำหรือวลีที่อยู่ในรูปแบบของทรัพยากร

เทคนิคการทำเหมืองความคิดเห็นหรือการวิเคราะห์ความคิดเห็นภาษาไทยที่มีอยู่ในปัจจุบันส่วนใหญ่มุ่งเน้นไปทางด้านการวิเคราะห์ความคิดเห็นที่ชัดเจน ซึ่งยังมีข้อจำกัดในการนำไปใช้ กล่าวคือ ความคิดเห็นที่นำมาวิเคราะห์ได้นั้นต้องมีคำที่ระบุข้อความความคิดเห็น (polar word) ที่ชัดเจน

แนวทางหนึ่งที่เป็นไปได้ในการแก้ข้อจำกัด ข้างต้นคือการวิเคราะห์ความคิดเห็นแบบปริยายที่อาศัยการสกัดคำหรือวลีที่แสดงคุณลักษณะของวัตถุที่สามารถสะท้อนความคิดเห็นแบบปริยาย ที่ไม่มีคำที่ระบุข้อความความคิดเห็นที่ชัดเจนโดยอาศัยวิธีการอ้างอิงจากคลังคำศัพท์ (corpus-based approach) ร่วมกับการพิจารณาบริบทในบริเวณใกล้เคียงกับคุณลักษณะนั้นภายในประโยคเพื่อนำไปสู่การระบุข้อความความคิดเห็นในที่สุด ดังอธิบายรายละเอียดในหัวข้อถัดไป

แนวทางในการวิเคราะห์ความคิดเห็นแบบปริยาย

จากการศึกษาทฤษฎีและวรรณกรรมที่เกี่ยวข้องกับการวิเคราะห์เหมืองความคิดเห็นทำให้สามารถกำหนดแนวทางในการวิเคราะห์ความคิดเห็นภาษาไทย แบบปริยายที่อาศัยการสกัดคำหรือวลีที่แสดงคุณลักษณะของวัตถุที่สามารถสะท้อนความรู้สึกแบบปริยาย โดยอาศัยวิธีการอ้างอิงจากคลังคำศัพท์ร่วมกับการพิจารณาบริบทในบริเวณใกล้เคียงกับคุณลักษณะนั้นภายในประโยคเพื่อนำไปสู่การระบุข้อความความคิดเห็น ซึ่งสามารถแสดงได้ด้วยกรอบแนวทางในการวิเคราะห์ความคิดเห็นแบบปริยายดังรูปที่ 1 ซึ่งแบ่งออกเป็น 3 มอดูล (modules)



รูปที่ 1 กรอบแนวทางในการวิเคราะห์ความคิดเห็นแบบปริยาย

1. **มอดูลการสร้างคลังความรู้ทางภาษา (Knowledge Corpus Construction Module)** เป็นส่วนของการสร้างคลังความรู้ทางภาษาที่สร้างขึ้นโดยผู้เชี่ยวชาญเพื่อนำไปใช้ในการสร้างคลังคำศัพท์ (lexicon corpus) ซึ่งประกอบด้วย

1.1 นิพจน์ระบุนาม (named entity) คือ นิพจน์ที่ทำหน้าที่ระบุชี้เฉพาะถึงสิ่งใด ๆ ที่สนใจ เช่น ชื่อบุคคล, ชื่อองค์กร, ชื่อสถานที่และชื่อสินค้าหรือชื่อบริการ

1.2 คำกริยาแบบเฉพาะเจาะจง (specific verbs) คือ คำกริยาที่มีการกำหนดหน้าที่แบบเฉพาะเจาะจง เช่น คำกริยาที่เกี่ยวกับการแสดงความรู้สึกหรืออารมณ์ และคำกริยาที่แสดงประสาทสัมผัสและการรับรู้

1.3 แบบอย่าง (patterns) คือ ชุดของกฎของความคิดเห็น (rules of opinions) แบบปริยาย ที่ใช้ในการกำหนดข้อความความคิดเห็น ซึ่งแสดงในรูปแบบของการจัดเรียงคำหรือวลีในประโยคภาษาไทยเพื่อนำไปใช้ในมอดูลการวิเคราะห์แบบอย่างความคิดเห็น

2. **มอดูลการเตรียมข้อมูล (Preprocessing)** เป็นส่วนการนำความคิดเห็นภาษาไทยจากแหล่งต่างๆ เช่น สื่อสังคมออนไลน์ ฟอรัม เว็บบอร์ด ที่รวบรวมไว้ในคลังข้อความแสดงความคิดเห็น (opinion corpus) และคลังความรู้ทางภาษามาผ่านกระบวนการเตรียมข้อมูลให้อยู่ในรูปแบบที่เหมาะสมเพื่อให้คอมพิวเตอร์สามารถนำไปประมวลผล มอดูลนี้ประกอบด้วย

2.1 การตัดคำ เป็นขั้นตอนของการนำข้อความแสดงความคิดเห็น ภาษาไทยที่เขียนเป็นข้อความต่อเนื่องกันมาทำการตัดเป็นคำ (word segmentation) เพื่อให้สามารถทำการวิเคราะห์ในระดับของคำได้

2.2 การกำหนดหน้าที่ของคำ ภายหลังจากการตัดคำในหัวข้อที่ 2.1 คำที่ตัดได้จะถูกนำมาทำการกำหนดหน้าที่โดยใช้คลังความรู้ทางภาษาที่ได้จากมอดูลในหัวข้อที่ 1 ร่วมกับการกำกับคำตามหน้าที่ของคำ (part-of-speech tagging) ซึ่งหน้าที่ของคำจะมีผลต่อความหมายของคำและการหาโครงสร้างของประโยค จากนั้นทำการสร้างเป็นรายการคำและจัดเก็บลงคลังคำศัพท์

3. **มอดูลการวิเคราะห์แบบอย่างความคิดเห็น** เป็นขั้นตอนของการสร้างโมเดลด้วยเทคนิคการเรียนรู้ของเครื่อง (machine learning technique) เช่น ซัพพอร์ตเวกเตอร์แมชชีน (support vector machine) ในการสร้างกฎเพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของข้อมูลที่ยังไม่เกิดขึ้นในการเปรียบเทียบระหว่างแบบอย่างประโยคในคลังข้อความแสดงความคิดเห็นที่ผ่านการสกัดคำศัพท์รวมถึงมีการกำกับหน้าที่ของคำไว้แล้วและแบบอย่างของประโยคจากคลังฐานความรู้ทางภาษาที่ได้จากผู้เชี่ยวชาญเพื่อทำการระบุข้อความความคิดเห็นต่อไป

นอกจากนี้โมเดลนี้ยังมีการประเมินค่าความแม่นยำ (precision) และค่าระลึก (recall) ของโมเดลการวิเคราะห์ข้อมูลที่ได้ และวัดค่า F-measure (ค่าการเฉลี่ยของค่าความแม่นยำและค่าระลึก) (B., Max, 2007: 173-176) เพื่อเป็นตัวบ่งชี้ความน่าเชื่อถือของโมเดลที่ได้ ดังสมการ

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$F - \text{measure} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

โดยที่ TP แทนจำนวนการทำนายที่ตรงกับข้อมูลจริงในคลาสที่กำลังพิจารณา FP แทนจำนวนการทำนายที่ทายผิดในคลาสที่กำลังพิจารณา และ FN แทนจำนวนการทำนายที่ทายผิดในคลาสที่ไม่ได้พิจารณา

สรุป

การทำเหมืองความคิดเห็นสามารถแบ่งตามการวิเคราะห์ความคิดเห็นออกเป็น 3 ระดับ ดังนี้คือ ระดับเอกสาร ระดับประโยค และที่ละเอียดและมีประสิทธิภาพที่สุด คือ ระดับคุณลักษณะของวัตถุ รูปแบบการวิเคราะห์ความคิดเห็นในระดับคุณลักษณะของวัตถุ สามารถแบ่งได้เป็น 2 รูปแบบ คือ การวิเคราะห์ความคิดเห็นที่ชัดเจนที่มีคำที่สะท้อนถึงความรู้สึกอย่างชัดเจน ปรากฏอยู่ในประโยค และการวิเคราะห์ความคิดเห็นแบบปริยายที่อาจไม่มีคำที่สะท้อนถึงความรู้สึกอย่างชัดเจน ปรากฏอยู่ในประโยค ซึ่งเทคนิคการทำเหมืองความคิดเห็นหรือ

การวิเคราะห์ความคิดเห็นแบบปริยายสำหรับข้อความภาษาไทย ยังคงมีจำนวนน้อยเนื่องจากเป็นภาษาที่ไม่มีขอบเขตของประโยค การแบ่งประโยคหรือแบ่งคำภาษาไทยไม่มีรูปแบบที่ตายตัว ไม่มีโครงสร้างที่ชัดเจน ทำให้มีกระบวนการที่ยุ่งยากและซับซ้อนกว่าภาษาอังกฤษ

จากการศึกษางานวิจัยที่มีอยู่แนวทางหนึ่งในการวิเคราะห์ความคิดเห็นแบบปริยายโดยการสกัดคำหรือวลีที่แสดงคุณลักษณะของวัตถุที่สามารถสะท้อนความรู้สึกแบบปริยายที่ไม่มีคำที่ระบุข้อความที่ชัดเจน คือ การอ้างอิงคลังคำศัพท์ ร่วมกับการพิจารณาบริบทในบริเวณใกล้เคียงกับคุณลักษณะนั้น โดยการเปรียบเทียบกับแบบอย่างของการจัดเรียงคำหรือวลีในประโยค ตามกฎของความคิดเห็นแบบปริยายเพื่อนำไปสู่การระบุข้อความความคิดเห็น

เอกสารอ้างอิง

- ตรีเทพ อารังลักษณ์ และ จันทรเจ้า มงคลนาวิน. 2554. "การพัฒนาระบบสรุปทวิจรรย์สินค้าภาษาไทยโดยผู้บริโภคแบบอัตโนมัติ." *จุฬาลงกรณ์ธุรกิจปริทัศน์*, 33, 128: 40-62.
- บุญเสริม กิจศิริกุล. 2546. *อัลกอริทึมการทำเหมืองข้อมูล*. กรุงเทพฯ: ภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย.
- วรัญญา วรณศรี. 2553. "ระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรม." สืบค้นเมื่อวันที่ 19 พฤศจิกายน 2553, จาก <http://thailang.nectec.or.th/halloffame/images/stories/best/download/13p33c001.pdf>
- Bramer, M. 2007. *Principles of Data Mining*. London: Springer. pp.173-176.
- Haruechaiyasak, C. et al. 2010. "Constructing Thai Opinion Mining Resource: A Case Study on Hotel Reviews." *Proc. of the Eighth Workshop on Asian Language Resources*, pp. 64-71.
- Hu, M., and Liu, B. 2004. "Mining Opinion Features in Customer Reviews." In *Proceedings of Nineteenth National Conference on Artificial Intelligence*.
- 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*, pp. 168-177.
- Kim, S. and E. Hovy. 2004. "Determining the Sentiment of Opinions." In *Proc. of the Intl. Conf. on Computational Linguistics (COLING'04)*, pp.1367-1373.
- Kongthon, A., Angkawattanawit, N. Sangkeetrakarn, C. Palingoon, P. and Haruechaiyasak, C. 2010. "Using an Opinion Mining Approach to Exploit Web Content in Order to Improve Customer Relationship Management." *Proceedings of Technology Management for Global Economic Growth*. Retrieved August 10, 2014 from IEEE Xplore Digital Library.
- Kongthon, A. et al. 2011. "HotelOpinion: An Opinion Mining System on Hotel Reviews in Thailand." *Technology Management in the Energy Smart World (PICMET), 2011 Proceedings of PICMET '11*. July 31 2011- 4 August 2011.
- Liu, B. 2012. *Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies*. Toronto: Morgan & Claypool.
- Oelke, D., et al. 2009. "Visual Opinion Analysis of Customer Feedback Data." *IEEE Symposium on Visual Analytics Science and Technology*, Atlantic City, New Jersey, USA.

- Pang, B., L. Lee, and S. Vaithyanathan. 2002. "Thumbs up? Sentiment Classification Using Machine Learning Techniques." In *Proc. of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, pp. 79-86.
- Popescu, A., and Etzioni, O. 2005. "Extracting Product Features and Opinions from Reviews." *Proceeding HLT '05 Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*: pp. 339-346.
- Qiu, G. et al. 2009. "Expanding Domain Sentiment Lexicon through Double Propagation." In *Proceedings of IJCAI 2009*.
- Riloff, E., Wiebe, J., and Wilson, T. 2003. "Learning subjective nouns using extraction pattern bootstrapping." *Proceedings of CoNLL 2003*.
- Sriphaew, K., Hiroya, T. and Manabu, O. 2009. "Sentiment Analysis for Thai Natural Language Processing." *Proceedings of the 2nd Thailand-Japan International Academic Conference, TJIA 2009*. Kyoto, Japan: pp. 123-124.
- Turney, P. 2002. "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews." In *Proc. of the Meeting of the Association for Computational Linguistics (ACL'02)*, pp. 417-424.
- Zhail, Y. et al. 2010. "Extracting Opinion Features in Sentiment Patterns." *International Conference on Information, Networking and Automation (ICINA)*.
- Zhang, L. and Liu, B. 2010. "Extracting and Ranking Product Features in Opinion Documents." *Proceedings of the 23rd International Conference on Computational Linguistics (COLING-2010)*.
- Zhang, L. and Liu, B. 2011. "Identifying Noun Product Features that Imply Opinions." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: shortpapers*: pp. 575-580.
- Zhang, L. and Liu, B. 2011. "Extracting Resource Terms for Sentiment Analysis." *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP-2011)*.



>> มาสวีย์ มาศดิตรชิต

สำเร็จการศึกษา ปริญญาโท (วท.ม.) สาขาเทคโนโลยีสารสนเทศ พ.ศ. 2541 สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ปริญญาตรี (วท.บ.) สาขาสถิติ พ.ศ. 2539 มหาวิทยาลัยศิลปากร
ปัจจุบันดำรงตำแหน่งหัวหน้าสาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยศรีปทุม