

# วิจัยเชิงสำรวจเพื่อศึกษาเทคนิคการวิเคราะห์ความรู้สึก

## ในการประมวลผลเนื้อหาบนสื่อสังคมออนไลน์

### A Survey of Sentiment Analysis Techniques on Social Media Content

กฤษฎาภรณ์ ปัญญา (Kritsayapon Panya), วรณนิสา ศรีทองคง (Wanisa Srithongkhong)

ศศิพิมพ์ ประสาทศรี (Sasipim Prasartsri), อภิวัฒน์ วัฒนะสุระ (Apiwat Wattanasura)

สาขาเทคโนโลยีสารสนเทศ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

Krisayaporn.p@kkumail.com, S.wannisa@kkumail.com, sasipimp@kkumail.com, apiwat.wat@kkumail.com

#### บทคัดย่อ

การวิจัยเรื่องวิจัยเชิงสำรวจเพื่อศึกษาเทคนิคการวิเคราะห์ความรู้สึกในการประมวลผลเนื้อหาบนสื่อสังคมออนไลน์มีวัตถุประสงค์เพื่อศึกษาเทคนิคการวิเคราะห์ความรู้สึกบนสื่อออนไลน์ในวิธีการต่างๆทางด้าน semantic ซึ่งมีการใช้เทคโนโลยีเป็นส่วนประกอบ โดยทำการเปรียบเทียบเพื่อหาข้อดีและข้อจำกัดในแต่ละวิธีเพื่อเป็นแนวคิดในการปรับปรุงเทคนิคของการวิเคราะห์ความรู้สึกในอนาคต

**คำสำคัญ:** การวิเคราะห์ความรู้สึก, สื่อสังคมออนไลน์, ความคิดเห็น

#### Abstract

A research study on the study of sensory analysis techniques for the processing of social media content aims to study the techniques of sensory analysis in online media in various semantic methods. Theory is a component. Comparison to

find the advantages and disadvantages of each method as a concept to improve the technique of the analysis of feelings in the future.

**Keyword:** Sentiment Analysis, Social Media, Opinion

#### 1. บทนำ

บนสื่อสังคมออนไลน์เต็มไปด้วยข้อมูลความคิดเห็นและความรู้สึกของผู้คนจำนวนมาก ซึ่งข้อมูลเหล่านี้ล้วนมีประโยชน์ในการวิเคราะห์เพื่อการวางแผนทางธุรกิจและการเข้าถึงความต้องการในด้านต่าง ๆ ในขณะที่มีข้อมูลจำนวนมากและความหลากหลาย แต่การใช้ประโยชน์จากข้อมูลต้องผ่านกระบวนการวิเคราะห์และสังเคราะห์ข้อมูล เพื่อให้ได้ข้อมูลสารสนเทศที่ตรงความต้องการของผู้ประกอบการที่สุด ทำให้ต้องมีกระบวนการวิเคราะห์ที่มีความแม่นยำและมีประสิทธิภาพเพื่อประโยชน์สูงสุดของข้อมูล ซึ่งปัจจุบันการวิเคราะห์ความรู้สึก (Sentiment Analysis) มีการเสนอกระบวนการและเทคนิคมากมาย เช่น การทำเหมืองข้อมูล (Data

Mining) , การใช้เทคโนโลยีสื่อความหมาย (semantic web). ซึ่งเป็นเทคโนโลยีที่ช่วยในการวิเคราะห์ความหมายและความสัมพันธ์ของคำ ซึ่งมีเทคนิคเช่น การสร้างโดเมนความรู้ (Ontology) และ เทคนิคการใช้คอมพิวเตอร์ในการเรียนรู้ (Machine Learning)

ในบทความนี้มีการนำเสนอเปรียบเทียบวิธีการในการวิเคราะห์ความรู้สึก จากความคิดเห็นบนสื่อสังคมออนไลน์ ในรูปแบบต่าง ๆ สำหรับศึกษาเทคนิคและกระบวนการ เพื่อใช้ในการตัดสินใจและหาแนวทางในการพัฒนาเพื่อให้ได้ประสิทธิภาพและความแม่นยำมากยิ่งขึ้น

## 2. ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง

### 2.1 การวิเคราะห์ความรู้สึก (Sentiment Analysis)

การวิเคราะห์ความรู้สึก(Sentiment Analysis) [1] เป็นการศึกษาเชิงคอมพิวเตอร์ในการแสดงความคิดเห็นทัศนคติและอารมณ์ของผู้ใช้ต่อเรื่องต่างๆ สามารถแสดงถึงบุคคลเหตุการณ์หรือหัวข้อ ซึ่งหัวข้อเหล่านี้ส่วนใหญ่ มีแนวโน้มที่จะครอบคลุมโดยบทวิจารณ์ ดังนั้นเป้าหมายของการวิเคราะห์ความรู้สึก(Sentiment Analysis) คือการวิเคราะห์ความคิดเห็นเพื่อบ่งบอกความรู้สึกที่ผู้ใช้แสดงออก และจัดประเภทข้อความความคิดเห็นของผู้ใช้ตามที่แสดงไว้

### 2.2 งานวิจัยที่เกี่ยวข้อง

Vivek Narayanan [2] ได้นำเสนอหัวข้อ Fast and accurate sentiment classification using an enhanced Naïve Bayes model เป็นงานวิจัยที่ศึกษาการจำแนกความรู้สึกที่เกี่ยวข้องกับภาพยนตร์ โดยใช้ข้อมูลมาจาก IMDB (Movie Reviews Database) ซึ่งเป็นแหล่งที่รวบรวมความ

คิดเห็นของภาพยนตร์ต่าง ๆ อีกทั้งงานวิจัยนี้ยังเพิ่มประสิทธิภาพการจำแนกความรู้สึกที่รวดเร็ว และถูกต้องโดยใช้อัลกอริทึม Naïve Bayes ในการจำแนกข้อมูล ซึ่งมีการแบ่ง Training ออกเป็น 25,000 ข้อความ และ Testing ออกเป็น 25,000 ข้อความ

L. A. Freitas [3] ได้นำเสนอหัวข้อ An Ontology based Sentiment Analysis for mobile products using tweets เป็นงานวิจัยที่ศึกษาการจำแนกความรู้สึกจากข้อความ Twitter ที่ลูกค้าได้แสดงความคิดเห็นที่เกี่ยวข้องกับโทรศัพท์ โดยใช้ข้อมูลของลูกค้าจำนวน 1,200 ข้อความ และใช้ SentiWordNet ในการหาค่าคะแนน Positive และ Negative ของคำ เมื่อวิเคราะห์คำเสร็จจะนำคะแนนมาเก็บไว้ในออนโทโลยีเพื่อใช้ในการสืบค้น

Deepak Kumar Yadav [4] ได้เสนอหัวข้อ Sentiment Analysis Based on Data Mining and Natural Language Processing เป็นงานวิจัยที่ศึกษาการวิเคราะห์ความรู้สึกที่รวบรวมข้อมูลมาจากเว็บไซต์ Amazon และ Flipkart โดยใช้หลากหลายเทคนิค ได้แก่ 1) Stop word เป็นการลบคำที่ไม่สำคัญทิ้งไปจากประโยค เช่น a, an, the, in, on เป็นต้น ซึ่งกลุ่มคำเหล่านี้เป็นคำที่ไม่เกี่ยวข้องกับการวิเคราะห์ความรู้สึก 2) Stemming เป็นการแปลงคำให้อยู่ในรูปแบบของรากศัพท์ (root) เช่น การตัด ing, ed, es ทิ้งไป 3) Stanford Parser ช่วยระบุ Parts Of Speech ของแต่ละคำ และสุดท้ายใช้ KNN Algorithm ในการจำแนกประเภทของคำวิจารณ์ระหว่างคำที่เป็น Positive กับ Negative

N.Kumar และคณะ [5] ได้ทำการศึกษาเรื่อง Sentiment dynamics in social media news channels วิเคราะห์ว่าเปรียบเทียบความเชื่อมั่นของโพสต์ข่าวสื่อสังคมออนไลน์ทางโทรทัศน์วิทยุและสื่อสิ่งพิมพ์วิเคราะห์ปฏิกิริยาของผู้ใช้และความคิดเห็นเกี่ยวกับโพสต์ข่าวสารที่มีความรู้สึกต่างกัน ซึ่งชุดข้อมูลที่สกัดจากหน้า Facebook ของ 5 ช่องข่าวยอดนิยม ได้แก่ CNN, Fox News, The Economist, NYT และ NPR ชุดข้อมูลมีข่าว 0.15 ล้านฉบับ และปฏิกิริยาผู้ใช้ 1.13 พันล้านคน ผลการทดลองแสดงให้เห็นว่าความรู้สึกของความคิดเห็นของผู้ใช้มีความสัมพันธ์กันอย่างมากกับความเชื่อมั่นของข่าวสารและประเภทของแหล่งข้อมูล การศึกษาของเรายังแสดงให้เห็นถึงความแตกต่างระหว่างช่องทางสื่อสังคมออนไลน์ของแหล่งข่าวต่างๆ

Hansi De Silva [6] ได้ทำการศึกษาเรื่อง Social media based personalized advertisement engine บทความนี้รายงานหลักฐานว่า AdSeeker เป็นเครื่องมือการกำหนดค่าตามความชอบของผู้ใช้ตาม Social Media เป็นวิธีแก้ปัญหาที่มีประสิทธิภาพในการปรับปรุงมูลค่าทางธุรกิจของการตลาดและการโฆษณา Adseeker เป็นระบบที่สร้างขึ้นโดยใช้การแมป ontological และการวิเคราะห์ความหมายของเนื้อหาโซเชียลมีเดียเพื่อให้ตรงกับข้อมูลส่วนบุคคล การระบุลำดับชั้นความสัมพันธ์ส่วนบุคคลและวิธีการ ontological เพื่อการจำแนกประเภทโฆษณาช่วยในการระบุการโฆษณาที่เหมาะสมที่สุดสำหรับผู้ใช้แต่ละราย AdSeeker ใช้ทวิตที่โพสต์โดยผู้ใช้เพื่อจับ

Gloria J.Kang และคณะ [7] ได้ทำการศึกษาเรื่อง Semantic network analysis of vaccine sentiment in online social media เพื่อตรวจสอบความเชื่อมั่นของวัคซีนในปัจจุบันเกี่ยวกับสื่อทางสังคมโดยการสร้างและวิเคราะห์เครือข่ายความหมายของข้อมูลวัคซีนจากเว็บไซต์ที่ใช้งานร่วมกันของผู้ใช้ Twitter ในสหรัฐอเมริกา และเพื่อช่วยในการสื่อสารเรื่องสาธารณสุขกับวัคซีน มีการรวบรวมทวิตทั้งหมด 26,389 รายการระหว่างวันที่ 16 เมษายน 2015 และวันที่ 29 พฤษภาคม 2015 ซึ่งเราได้รับลิงก์เว็บที่ไม่ซ้ำกัน 8416 รายการ

ทั้งนี้ยังมีข้อจำกัดคือ ไม่มีการประเมินค่าความเชื่อมั่น, ภาษาคลุ่มเครือเช่น คำสแลงคำพังเพยและบทกวี

Karen A.Monsen และคณะ [8] ได้ทำการศึกษาเรื่อง Use of the Omaha System for ontology-based text mining to discover meaning within CaringBridge social media journals เป็นการประเมินความเป็นไปได้ในการใช้คำศัพท์เกี่ยวกับอินเทอร์เฟซและ ontology (Omaha System) เป็นรูปแบบข้อมูลการทำเหมืองข้อมูลข้อความปรับกระบวนการในการพัฒนาเนื้อหา CaringBridge และขยายการใช้วิธีนี้ไปยังรายการวารสารและคำศัพท์ที่ผู้บริโภคร่างขึ้น

Teng Niu และคณะ [9] ได้ทำการศึกษาเรื่อง Sentiment Analysis on Multi-View Social Data เป็นการแนะนำชุดข้อมูลใหม่ที่เรียกว่า MVSA ซึ่งประกอบด้วยทวิตหลายมุมมองสำหรับการวิเคราะห์ความเชื่อมั่น ผลของเราสามารถใช้ประโยชน์ได้เป็นพื้นฐานนอกเหนือจากการวิเคราะห์ความเชื่อมั่นแล้วชุดข้อมูลของเรายังสามารถใช้ใน

การตรวจสอบปัญหาที่น่าสนใจอื่น ๆ ที่เปิดเช่นการตรวจสอบ tweets นอกจากนี้ยังได้แสดงให้เห็นว่ามีป้ายข้อความที่ไม่สอดคล้องกันหลายอย่างระหว่างข้อความมุมมองและภาพในทวีตที่เก็บรวบรวม

Dazhen Lin และคณะ [10] ได้ทำการศึกษาเรื่อง Multi-modality weakly labeled sentiment learning based on Explicit Emotion Signal for Chinese microblog เป็นการคาดการณ์เกี่ยวกับความเชื่อมั่นข้ามสื่อ โดยใช้ Visual Sentiment Ontology (VSO), Emotion words ฯลฯ ต้องใช้ตัวอย่างที่มีข้อความจำนวนมากเพื่อเรียนรู้คุณสมบัติระดับกลาง โดยยังไม่สามารถใช้งานได้กับแอปพลิเคชันเครือข่ายสังคมออนไลน์แบบเรียลไทม์ วิธีนี้ใช้ตัวอย่างอย่างน้อยสองสามคนที่มีข้อความกำกับเพื่อให้ได้ผลการปฏิบัติงานเหมือนกันซึ่งได้จากวิธีการทำนายความเชื่อมั่นของเครื่องเรียนแบบดั้งเดิม

Pratik Thakor And Dr. Sreela Sasib [11] ได้ทำการศึกษาเรื่อง Ontology-based Sentiment Analysis Process for Social Media Content วิเคราะห์เนื้อหาโซเชียลมีเดีย (OSAPS) ด้วยความรู้สึกเชิงลบเนื้อหาโซเชียลมีเดียถูกแยกออกจากข้อความ twitter โดยอัตโนมัติ กระบวนการ ontology-based

Samir Tartir And Ibrahim Abdul-Nabi [12] ได้ทำการศึกษาเรื่อง Semantic Sentiment Analysis in Arabic Social Media เป็นการ วิเคราะห์ความคิดเห็นมีวัตถุประสงค์เพื่อกำหนดทัศนคติของกลุ่มคนที่ใช้แพลตฟอร์มโซเชียลมีเดียอย่างน้อยหนึ่งรายการในหัวข้อที่เฉพาะเจาะจง ในบทความนี้เราเสนอแนวทางเชิงความหมาย

เพื่อค้นหาทัศนคติของผู้ใช้และข้อมูลเชิงลึกทางธุรกิจจากสื่อสังคมออนไลน์ในภาษาอาหรับทั้งในแบบมาตรฐานและภาษาถิ่น

R.Nithish, S.Sabarish, Dr. A.Askarunisa [13] ได้ทำการศึกษาเรื่อง An Ontology based Sentiment Analysis for mobile products using tweets วิเคราะห์ผลิตภัณฑ์ของตนเข้าถึงโดยใช้แบบสำรวจความคิดเห็นและแบบสอบถามอื่น ๆ โดยใช้แบบฟอร์มออนไลน์ข้อความที่ไม่มีโครงสร้างทั้งหมดเหล่านี้และข้อมูลปริมาณมากต้องเป็นวิเคราะห์ เพื่อทำความเข้าใจกับบริบทและตัดสินใจเกี่ยวกับธุรกิจ

### 3. ขั้นตอนการดำเนินงานวิจัย

งานวิจัยนี้เป็นการค้นหางานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความรู้สึกมาวิเคราะห์

### 4. ผลการดำเนินงาน

ตารางที่ 1 : ตารางเปรียบเทียบเทคนิควิธีการของแต่ละงานวิจัย

บทความ	เทคนิค/วิธีการ	ผลการทดลอง (ร้อยละ)
[2]	Naïve Bayes	88.80
[3]	ใช้ SentiWordNet	70.00
[4]	1) Stop word 2) Stemming 3) Stanford Parser ช่วยระบุ Parts Of Speech	-
[5]	VADER เป็นฐานคำศัพท์ที่เหมาะสมสำหรับการวิเคราะห์ความเชื่อมั่นของเนื้อหาที่มี	CNN 97.00 Fox News 98.00 Economist 95.00

	ต้นกำเนิดในสื่อสังคมออนไลน์	NYT 97.00 NPR 93.00
[6]	เครื่องมือนำเสนอ JAVA JAX-RX บริการเว็บ API Restful API, Apache Jena และค้นหาด้วย ontology, hibernate, MySQL Server, Naive Bayes และ NLTK	88.00
[7]	ใช้ sentiment networks โดยมี NetworkX & iGraph ในการสร้างและ วิเคราะห์เครือข่าย	-
[8]	การทำเหมืองข้อความที่มี โครงสร้างแบบ ontology โดยใช้ KONTAKT 3	-
[9]	ทดลองเกี่ยวกับการจัด หมวดหมู่ข้อความความรู้สึก โดยการใช้ข้อมูลที่ระบุว่า เป็นบวกและลบในชุดข้อมูล ที่สร้างขึ้นของเรา ชุดข้อมูล ถูกสุ่มแบ่งเป็นการฝึกอบรม และการทดสอบ 50%-50%	T-V-Early 75.20 T-V-Late 72.50 M-DBM 74.70
[10]	Visual Sentiment Ontology (VSO), Emotion words	60.10
[11]	ใช้ซอฟต์แวร์ GATE [11] การวิเคราะห์อัตโนมัติข้อมูล tweet	-
[12]	Naive Bayes	75.00

[13]	การสกัดตามคุณลักษณะ ของทวีต	Op_Score (f) = -2 ถ้าความคิดเห็น (f) <= -0.5 (แย่ที่สุด); Op_Score (f) = -1 ถ้า -0.5 < opinion (f) < 0 (ไม่ดี) Op_Score (f) = 0 ถ้า ความคิดเห็น (f) = 0 (เป็นกลาง); Op_Score (f) = + 1 ถ้า 0 <ความคิดเห็น (f) <0.5 (ดี) Op_Score (f) = +2 if opinion (f) > 0.5 (excellent)
------	--------------------------------	--

## 5. สรุปผล

งานวิจัยนี้เป็นการเทคนิคการวิเคราะห์ความรู้สึก เพื่อประมวลผลเนื้อหาสื่อสังคมออนไลน์ โดยผู้วิจัยได้ ทำการศึกษางานวิจัยที่เกี่ยวข้องเพื่อค้นหาเทคนิคที่เหมาะสมและดีมากที่สุด พบว่าหากมีการตัดคำและมีการคิด ค่าคะแนนจากหน้าที่ของคำได้ถูกต้อง จะให้ค่าประสิทธิภาพ ที่ดี ดังนั้นหากต้องการทำเกี่ยวกับวิเคราะห์ความรู้สึกควรใช้ วิธีนี้ร่วมด้วยในการวิเคราะห์

## 6. เอกสารอ้างอิง

- [1] Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*. 2014;5(4):1093-113.
- [2] Vivek Narayanan (2013), Fast and accurate sentiment classification using an enhanced Naive Bayes model, *Intelligent Data Engineering and Automated Learning IDEAL 2013 Lecture Notes in Computer Science Volume 8206*, 2013, pp 194-201.
- [3] L. A. Freitas and R. Vieira, An Ontology based Sentiment Analysis for mobile products using tweets. *22nd International Conference on World Wide Web*, 2013, pp. 367-370: ACM.
- [4] Yadav, Deepak Kumar, and Sampada Vishwas Massey. "Sentiment Analysis Based on Data Mining and Natural Language Processing."
- [5] Kumar N, Nagalla R, Marwah T, Singh M. Sentiment dynamics in social media news channels. *Online Social Networks and Media*. 2018;8:42-54.
- [6] Silva HD, Jayasinghe P, Perera A, Pramudith S, Kasthurirathna D, editors. *Social media based personalized advertisement engine*. 2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA); 2017 6-8 Dec. 2017.
- [7] Kang GJ, Ewing-Nelson SR, Mackey L, Schlitt JT, Marathe A, Abbas KM, et al. Semantic network analysis of vaccine sentiment in online social media. *Vaccine*. 2017;35(29):3621-38.
- [8] Monsen KA, Maganti S, Giaquinto RA, Mathiason MA, Bjarnadottir RI, Kreitzer MJ. Use of the Omaha System for ontology-based text mining to discover meaning within CaringBridge social media journals. *Kontakt*. 2018;20(3):e210-e6.
- [9] Teng Niu SZ, Lei Pang, and Abdulmotaleb El Saddik. *Sentiment Analysis on Multi-View Social Data*. 2016. p. 15-27.
- [10] Lin D, Li L, Cao D, Lv Y, Ke X. Multi-modality weakly labeled sentiment learning based on Explicit Emotion Signal for Chinese microblog. *Neurocomputing*. 2018;272:258-69.
- [11] S. Tartir and I. Abdul-Nabi, "Semantic Sentiment Analysis in Arabic Social Media," *Journal of King Saud University - Computer and Information Sciences*, vol. 29, no. 2, pp. 229-233, 2017.
- [12] P. Thakor and S. Sasi, "Ontology-based Sentiment Analysis Process for Social Media Content," *Procedia Computer Science*, vol. 53, pp. 199-207, 2015.
- [13] R. Nithish, S. Sabarish, M. N. Kishen, A. M. Abirami, and A. Askarunisa, "An ontology based sentiment analysis for mobile products using tweets," in *2013 Fifth International Conference on Advanced Computing (ICoAC)*, 2013, pp. 342-347.