

# แนวคิดในการสร้างและการใช้ประโยชน์ ของ Sentiment Corpus

กัลยา ทองเอียด<sup>1</sup> และ รศ.ดร. พรฤดี เนติโสภาคกุล<sup>2</sup>

<sup>1</sup>คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ

<sup>2</sup>คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ

Emails: kanlaya.satin@gmail.com, ponrudee@it.kmitl.ac.th

## บทคัดย่อ

การแสดงความคิดเห็นรวมถึงอารมณ์ความรู้สึกในรูปแบบข้อความ ของผู้ใช้งานบนอินเทอร์เน็ตเกิดขึ้นตลอดเวลา ซึ่งหากข้อความมีจำนวนมากจนมนุษย์ไม่สามารถวิเคราะห์ได้ในเวลาอันจำกัด ก็ไม่สามารถนำข้อความเหล่านั้นมาใช้ประโยชน์ได้ งานด้านการวิเคราะห์ความคิดเห็นและการวิเคราะห์อารมณ์จากข้อความจึงมีความสำคัญมาก ซึ่งการวิเคราะห์ดังกล่าวมีหลายแนวทาง แต่แนวทางที่ที่จะต้องประกอบด้วยทรัพยากรคำศัพท์ที่บ่งบอกข้อความคิดเห็น บทความนี้จึงได้ทบทวนวรรณกรรมเกี่ยวกับการสร้างเครื่องมือที่ช่วยในการวิเคราะห์ความรู้สึกจากข้อความแบบอัตโนมัติ โดยการรวบรวมคำศัพท์ที่ใช้ในการวิเคราะห์ข้อความ ที่เกี่ยวข้องกับอารมณ์ ความรู้สึกของมนุษย์ที่สร้างขึ้นโดยใช้เทคนิค AI และเว็บเชิงความหมาย และสร้างเป็นทรัพยากร ที่เรียกว่า SenticNet ซึ่งเป็นทรัพยากรที่ใช้ในการวิเคราะห์ความรู้สึกจากข้อความที่เปิดให้ใช้งานแบบสาธารณะ ผู้ที่ต้องการใช้งานสามารถดาวน์โหลดไปเชื่อมต่อกับโปรแกรมที่ใช้แยกคำจากข้อความที่เป็นภาษาธรรมชาติได้ โดย SenticNet จะแสดงค่าความรู้สึก ค่าชั่วอารมณ์ เพื่อบอกว่าข้อความเหล่านั้นให้ความรู้สึกเป็นอย่างไร

**คำสำคัญ** –การทำเหมืองความคิดเห็น; การวิเคราะห์อารมณ์ความรู้สึก; ทรัพยากรด้านอารมณ์; เซนติเน็ต

**Keyword**– Opinion Mining; Sentiment Analysis; Sentiment Resource; SenticNet;

## 1. บทนำ

ข้อความหรือความคิดเห็นที่ปรากฏบนสื่อสังคมออนไลน์ อาทิเช่น เว็บเพจ เฟสบุ๊ค ทวิตเตอร์ ฟันทิป หรือสื่อโซเชียลอื่นๆ เป็นต้น เป็นสิ่งที่มีประโยชน์ต่อผู้เกี่ยวข้อง อาทิเช่น นักธุรกิจ ผู้บริหาร ผู้กำหนดนโยบายสาธารณะ และผู้อ่านทั่วไป อย่างไรก็ตาม ปริมาณข้อความมีขนาดใหญ่และเพิ่มขึ้นอย่างรวดเร็ว การวิเคราะห์ข้อความเหล่านี้โดยมนุษย์ จะทำได้ไม่ทันการณ์ ดังนั้น จึงมีแนวคิดที่จะวิเคราะห์ข้อความโดยอัตโนมัติ เพื่อให้สามารถ

วิเคราะห์ข้อความได้อย่างรวดเร็ว

ปัจจุบันมีเครื่องมือช่วยวิเคราะห์ข้อความ อาทิเช่น keyword spotting เครื่องมือทางสถิติ เป็นต้น ซึ่งสามารถวิเคราะห์ข้อความได้ในระดับหน้าเอกสารเท่านั้น แต่ไม่สามารถวิเคราะห์ในระดับคำหรือระดับประโยคได้ และเนื่องด้วยแหล่งทรัพยากรที่ใช้ในการวิเคราะห์ข้อความมีไม่เพียงพอต่อข้อความจำนวนมาก ด้วยเหตุนี้จึงเกิดการพัฒนาศenticNet ขึ้นมา ซึ่งเป็นทรัพยากรที่

รวบรวมข้อมูลคำศัพท์ที่ใช้ในการวิเคราะห์ข้อความ ที่เกี่ยวข้องกับอารมณ์ ความรู้สึกของมนุษย์ที่สร้างขึ้นโดยใช้เทคนิค AI และเว็บเชิงความหมาย ที่เปิดให้ใช้งานแบบสาธารณะในรูปแบบสแตนด์อโลนและรูปแบบออนไลน์ หรือ API

ส่วนที่ 2 ของบทความนี้กล่าวถึงความเป็นมาของ SenticNet ในเรื่องของแนวคิดที่นำมาใช้เป็นความรู้ในการสร้าง ประวัติความเป็นมาของ SenticNet และการเปรียบเทียบ SenticNet แต่ละเวอร์ชัน ส่วนที่ 3 กล่าวถึงกระบวนการสร้าง SenticNet ส่วนที่ 4 กล่าวถึงการนำ SenticNet ไปใช้งาน และส่วนที่ 5 เป็นส่วนสรุป และข้อเสนอแนะ ที่สามารถนำไปปรับปรุงในอนาคตต่อไป

## 2. ความเป็นมาของ SenticNet

### 2.1 นิยามคำศัพท์

คลังภาษา (Language Corpus) คือ คลังข้อมูลทางด้านภาษาที่มาจากเอกสาร หนังสือ หนังสือพิมพ์ หรือจากแหล่งข้อมูลทางด้านภาษาอื่นๆ มาจัดทำอยู่ในรูปแบบอิเล็กทรอนิกส์ ให้คอมพิวเตอร์ประมวลผลได้ โดยมีการกำกับข้อมูลเพิ่มเติม เช่น ประเภทของคำ หรือข้อมูลอื่นๆ ที่ต้องการใช้งานไว้แล้ว ซึ่งอาจจะทำโดยมนุษย์หรือทำโดยอัตโนมัติก็ได้ เช่น Brown corpus ที่มีการกำกับ POS ไว้แล้ว ส่วน Sentiment Corpus คือ คลังภาษาที่มีการกำกับค่าอารมณ์ (Sentiment) ไว้แล้ว ซึ่งรวบรวมมาจากการกำกับของฝูงชน (Crowd Sourcing) เช่นโครงการ Open Mind Common Sense Project ที่นำมาใช้ในการสร้าง ConceptNet เป็นต้น

ทรัพยากรทางด้านภาษา (Language Resource) คือ ทรัพยากรที่ใช้ในการประมวลผลภาษาด้วยคอมพิวเตอร์ เช่น WordNet, Online Dictionary โดย Language Corpus และ SenticNet ถือเป็นทรัพยากรทางด้านภาษาอีกรูปแบบหนึ่งที่จะนำมากล่าวถึงในบทความนี้

### การทำเหมืองความคิดเห็น (Opinion Mining)

คือ การรวบรวมความคิดเห็นจากหลายๆข้อความ ในเรื่องใดเรื่องหนึ่งโดยเฉพาะ เพื่อนำมาวิเคราะห์ความคิดเห็น มักจะวิเคราะห์เป็นเชิงบวก เชิงลบ หรือเชิงความเป็นจริงเท่านั้น

### การวิเคราะห์ความรู้สึก (Sentiment analysis)

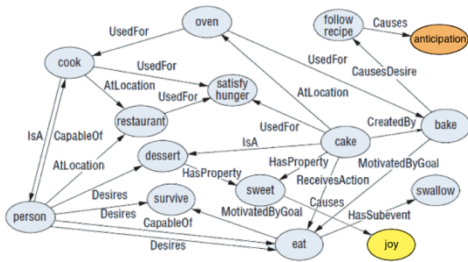
คือ กระบวนการเชิงคำนวณเพื่อกำกับข้อความหรือส่วนของข้อความที่แสดงความคิดเห็นด้วยทัศนคติของผู้เขียนข้อความนั้นๆ ต่อสินค้าหรือหัวข้อหนึ่งๆ ซึ่งมักจะเป็นทัศนคติเชิงบวก เชิงลบ หรือเป็นกลาง อย่างไรก็ตาม ในการกำกับข้อความระดับคำ สามารถใช้ทรัพยากร SenticNet ในการแทนค่าอารมณ์ Sentic Value 4 มิติ คือ ความพึงพอใจ (Pleasantness) ความสนใจ (Attention) ความอ่อนไหว (Sensitivity) และ ความฉลาด (Aptitude)

SenticNet คือ ทรัพยากรสำหรับการทำเหมืองความคิดเห็น ที่สร้างขึ้นโดยใช้เทคนิค AI และเว็บเชิงความหมาย เพื่อหาข้อความความคิดเห็นของ Concept ต่างๆ โดยประมวลผลข้อความภาษาธรรมชาติในระดับเชิงความหมาย

WordNet Affect คือ คลังข้อมูลคำศัพท์ที่เกี่ยวข้องกับอารมณ์ และความรู้สึก

ConceptNet คือ การแทนข้อมูล Concept ที่รวบรวมได้จากอาสาสมัครจำนวนมากบนอินเทอร์เน็ต ตั้งแต่ปี พ.ศ.2543 เป็นต้นมา ในโครงการ Open Mind Common sense Project และเก็บรวบรวมใน Open Mind Corpus โดยจะแทน Concept เหล่านี้ด้วยกราฟ โดยมีโหนดแสดง Concept และมีเส้นที่กำกับความสัมพันธ์ของโหนดทั้งสอง เช่น IsA UsedFor เป็นต้น [1] [2]

COGBASE คือ ฐานความรู้ที่ประกอบด้วยความรู้ทั่วไป (Common Knowledge) และความรู้สามัญสำนึก (Common-sense Knowledge)



รูปที่ 1. แสดงเครือข่ายของคำบน ConceptNet[2]

AffectiveSpace คือ การแทนอารมณ์ ความรู้สึก ในคอมพิวเตอร์ ด้วยเวกเตอร์ หลายมิติ ซึ่งแต่ละมิติมี 2 ขั้ว คือ บวกและลบ ใช้ในการกำกับอารมณ์ ความรู้สึก ให้กับข้อความ [3]

Aligning คือ การนำทรัพยากรจากหลายๆแหล่ง มาจัดเรียงให้อยู่บนระนาบเดียวกัน ด้วยการนำมาซ้อนทับ กันเพื่อให้เกิดมิติ

Singular value decomposition (SVD) คือ การทำให้ Concept ที่แสดงอารมณ์เดียวกันหรือใกล้เคียง กัน มีแนวโน้มที่จะอยู่ในตำแหน่งใกล้เคียงกันบน

AffectiveSpace ด้วยวิธีการหมุนแกน

การลดมิติ (Dimensional Reduction) เนื่องจาก ในแต่ละ Concept มีองค์ประกอบจำนวนมากทำให้ Concept เป็นมิติขนาดใหญ่ ดังนั้น การลดมิติของ Concept สามารถทำได้โดยการเลือกองค์ประกอบ 100 องค์ประกอบที่สำคัญที่สุด เพื่อทำให้มิติของ Concept บนเวกเตอร์มีขนาดเล็กลง

CF-IOE คือ เทคนิคที่ใช้ในการระบุความหมาย ให้กับ Concept โดยการหา Concept ที่มีความสัมพันธ์ กันในเชิงความหมาย และระบุให้กับ Concept นั้นๆ [7]

Hourglass of Emotions and Sentic value คือ การจำแนกประเภทของอารมณ์ สร้างเป็นแบบจำลองทาง อารมณ์ โดย Robert Plutchiks ได้ศึกษาอารมณ์ของ มนุษย์ และได้จัดโครงสร้างใหม่เป็น 4 มิติ ที่เรียกรวมกัน ว่าค่าความรู้สึก แต่ละมิติเป็นอิสระจากกัน จะให้ค่าที่เป็น บวกและลบ โดยแต่ละมิติจะมีระดับของอารมณ์ความรู้สึก 6 ระดับ ดังแสดงในตารางที่ 1 [4]

ตารางที่ 1. แสดงอารมณ์ 6 ระดับ จากค่า Sentic 4 มิติ ตามนิยามของ Hourglass of Emotion

	Pleasantness	Attention	Sensitivity	Aptitude
3	ecstasy	vigilance	rage	admiration
2	joy	anticipation	anger	trust
1	serenity	interest	annoyance	acceptance
-1	pensiveness	distraction	apprehension	boredom
-2	sadness	surprise	fear	disgust
-3	grief	amazement	terror	loathing

## 2.2 ประวัติการสร้าง SenticNet

SenticNet ได้ริเริ่มขึ้นในปี พ.ศ. 2552 ที่มหาวิทยาลัย

สเตอร์ลิง ประเทศอังกฤษ โดย Erik Cambria และคณะ

[5] มีจุดมุ่งหมายเพื่อใช้ในงานประมวลผลข้อความที่มีความคิดเห็นต่างๆ ที่เกิดขึ้นในชีวิตประจำวัน โดยเริ่มจากการวิจัย ค้นคว้า และพัฒนาเป็น SenticNet ในเวอร์ชันต่างๆ ดังนี้

### 2.2.1 SenticNet 1

SenticNet1 สร้างขึ้นเมื่อปี พ.ศ.2553 โดยใช้แหล่งทรัพยากรจาก WordNet Affect และ ConceptNet ใน Open Mind Corpus โดยผ่านกระบวนการทำให้ Concept อยู่ในรูปของ AffectiveSpace เพื่อนำ Concept นั้นมาหาค่าชั่วอารมณ์ (Polarity) ที่มีชั่วเป็นค่าระหว่าง [-1,+1] เท่านั้น โดยคำนวณค่าชั่วอารมณ์จากค่าอารมณ์ (Sentic Values) 4 มิติของคำ ได้แก่ ค่า Pleasantness และ Aptitude ซึ่งจะถูกคิดเป็นความรู้สึกเชิงบวก ส่วนค่า Attention และ Sensitivity จะถูกคิดเป็นความรู้สึกเชิงลบ ซึ่งสามารถรวบรวมคำศัพท์ได้จำนวน 5,700 คำ นำมาเข้ารหัส ในรูปแบบ RDF/XML แล้วเก็บในฐานข้อมูลของ SenticNet และสร้างเป็น แอปพลิเคชันที่เปิดให้ใช้งานแบบสาธารณะ[6]

### 2.2.2 SenticNet 2

SenticNet2 สร้างขึ้นเมื่อปี พ.ศ.2555 เป็นการปรับปรุงจาก SenticNet1 โดยใช้แหล่งทรัพยากรจาก Live-Journal ซึ่งเป็นเว็บไซต์ที่มีการ tag หัวข้อ มีลักษณะคล้ายกับพันทิป สามารถรวบรวมคำศัพท์ได้จำนวน 14,000 คำ โดยแต่ละคำ จะถูกสร้างความสัมพันธ์ในเชิงความหมายและนำมาหาค่าความรู้สึก ซึ่งผลลัพธ์ที่ได้ คือ คำที่มีความหมายเดียวกันหรือใกล้เคียงกับคำที่ต้องการจำนวน 10 คำ และค่าความรู้สึก 4 คำ จากอารมณ์ 4 มิติ [7] ความแตกต่างระหว่าง SenticNet1 และ SenticNet2 คือ SenticNet2 เริ่มมีการกำกับอารมณ์ของแต่ละคำ แยกเป็น 4 มิติ ขณะที่ SenticNet1 แทนด้วยค่า Polarity ซึ่งอยู่ระหว่าง [-1,1] นอกจากนี้ ใน SenticNet2 ยังมีการระบุคำที่มีความหมายและอารมณ์ใกล้เคียงกันอีก 10 คำ

ด้วย

### 2.2.3 SenticNet 3

SenticNet3 สร้างขึ้นเมื่อปี พ.ศ.2557 โดยใช้ฐานความรู้ COGBASE ซึ่งจะแตกต่างจาก SenticNet เวอร์ชันก่อนหน้า ที่ใช้แหล่งทรัพยากรจาก Open Mind Corpus เท่านั้น ทำให้ SenticNet3 มีจำนวนคำศัพท์เพิ่มขึ้นเป็น 25,000 คำ และมีประสิทธิภาพในการวิเคราะห์ความรู้สึกที่มากกว่า [8] ดังตารางที่ 2

ตารางที่ 2. แสดงประสิทธิภาพในการวิเคราะห์ความรู้สึก จากความคิดเห็นของผู้ป่วย ด้วย SenticNet[8]

Category	SenticNet 1	SenticNet 2	SenticNet 3
Clinical service	59.12%	69.52%	78.06%
Communication	66.81%	76.35%	80.12%
food	67.95%	83.61%	85.94%
parking	63.02%	75.09%	79.42%
staff	58.37%	67.90%	76.19%
timeliness	57.98%	66.00%	75.98%

### 2.2.4 SenticNet 4

SenticNet4 สร้างขึ้นเมื่อปี พ.ศ.2559 ใช้ประโยชน์จากคำศัพท์เดิมของ SenticNet เวอร์ชันก่อนหน้า แต่เพิ่มเติมคือการนำคำศัพท์เหล่านั้นมาแยกเป็นคำนามและคำกริยา แล้วนำมารวมกันเพื่อให้เกิดเป็นความหมายใหม่ อาทิเช่น คำว่า “buy\_cake” เป็นการนำคำกริยา คือ “buy” และ คำนาม คือ “cake” มารวมกัน เป็นต้น และเก็บรวบรวมไว้ใน SenticNet ซึ่งสามารถรวบรวมคำศัพท์ได้จำนวน 50,000 คำ[2]

## 2.3 การเปรียบเทียบ SenticNet แต่ละ Version

SenticNet แต่ละเวอร์ชัน มีความแตกต่างกัน เนื่องจาก

การพัฒนาที่มีความต่อเนื่องกัน ความรู้ที่นำมาใช้ที่เพิ่มมากขึ้นทำให้จำนวนของคำศัพท์ที่ได้จากการรวบรวมความรู้มีมากขึ้นด้วยเช่นกัน ในบทความนี้ขอเปรียบเทียบ SenticNet แต่ละเวอร์ชัน ในเรื่องของปีที่สร้าง แหล่ง

ทรัพยากรที่นำมาใช้ ผลลัพธ์ที่ได้ และจำนวนคำศัพท์ที่รวบรวมได้ เพื่อให้ง่ายต่อความเข้าใจ จึงได้นำเสนอการเปรียบเทียบในรูปแบบตาราง ดังตารางที่ 3

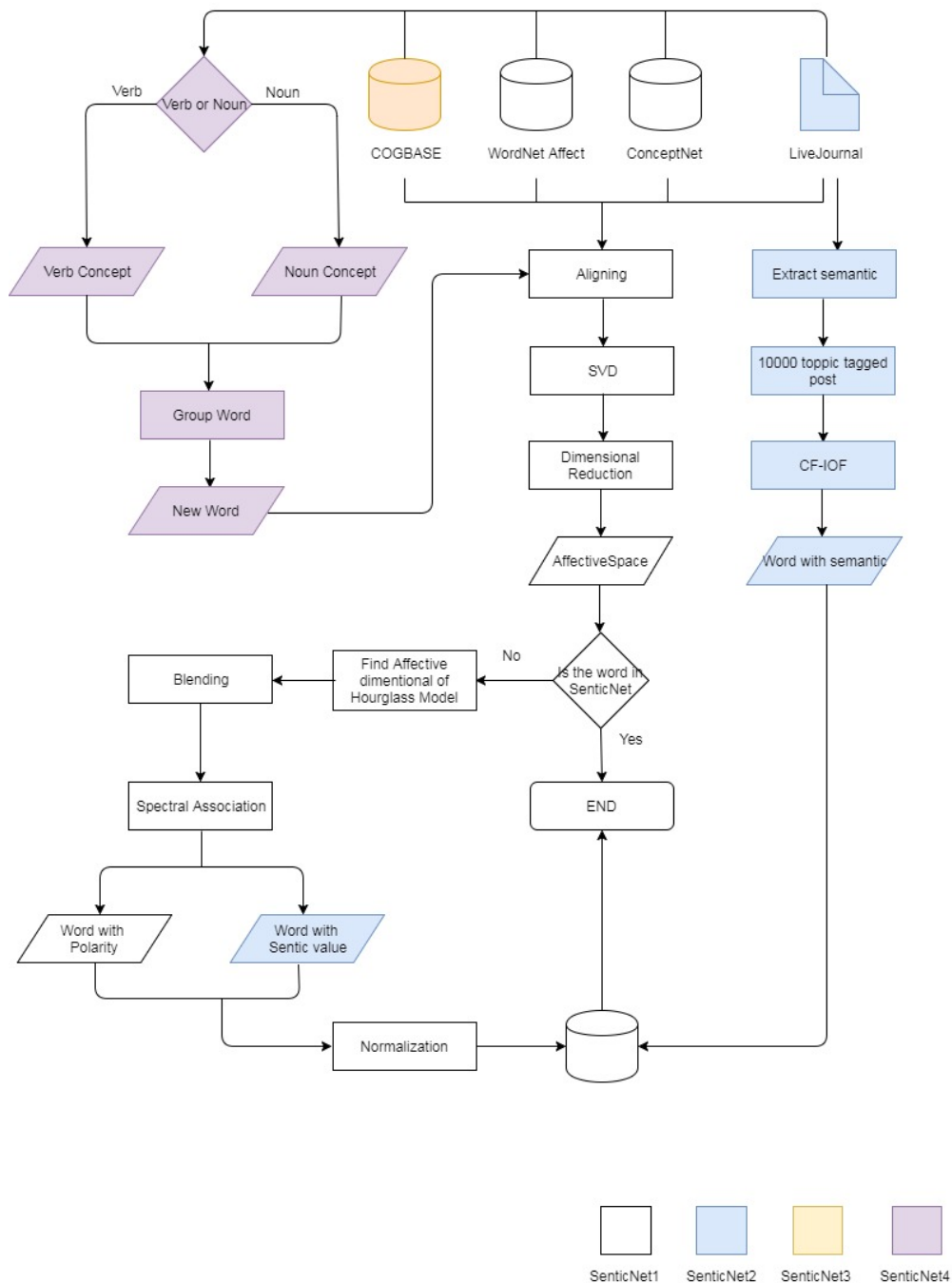
**ตารางที่ 3.** การเปรียบเทียบ SenticNet แต่ละเวอร์ชัน

เวอร์ชัน	ปีที่สร้าง	แหล่งทรัพยากร	ผลลัพธ์	คำศัพท์
SenticNet 1	พ.ศ.2553	WordNet Affect และ ConceptNet ใน Open mind corpus	คำซ้ำอารมณ์	5,700 คำ
SenticNet 2	พ.ศ.2555	Live Journal	- คำที่มีความหมายเดียวกัน 10 คำ - คำความรู้สึก	14,000 คำ
SenticNet 3	พ.ศ.2557	ฐานความรู้ COGBASE	- คำซ้ำอารมณ์ - คำที่มีความหมายเดียวกัน 10 คำ - คำความรู้สึก	25,000 คำ
Senticnet 4	พ.ศ.2559	ใช้คำศัพท์เดิมจาก SenticNet เวอร์ชันก่อนหน้า แต่เพิ่มเติมคือการนำคำศัพท์เหล่านั้นมาแยกเป็นคำนามและคำกริยา และนำคำที่แยกมารวมกันเพื่อให้เกิดเป็นคำใหม่ที่มีความหมาย	คำใหม่ที่มีความหมาย	50,000 คำ

### 3. กระบวนการสร้าง SenticNet

ภาพรวมของกระบวนการสร้าง SenticNet เริ่มจากการรวมคำศัพท์จากแหล่งทรัพยากรหลักตั้งแต่ 2 แหล่ง คือ WordNetAffect และConceptNet ใน SenticNet1 และเพิ่มเติมคำศัพท์จาก LiveJournal ใน SenticNet2 และเพิ่มคำศัพท์จากฐานข้อมูล COGBASE ใน SenticNet3 นอกจากนี้ยังมีกระบวนการสร้างคำใหม่จากคำนามและคำกริยาในทรัพยากรหลักใน SenticNet4 คำศัพท์ทั้งหมดจะถูกแทนอยู่ในรูปของ AffectiveSpace ที่เกิดจาก

กระบวนการ Aligning ในการจัดการทรัพยากร แล้วใช้ SVD ในการทำให้ Concept เกิดมิติ หลังจากนั้นนำมาผ่านกระบวนการลดมิติ จะทำให้มิติของ Concept บนเวกเตอร์มีขนาดเล็กลง นำเวกเตอร์ที่ได้จาก AffectiveSpace มาหาข้อสรุปของค่าความรู้สึกและค่าซ้ำอารมณ์ โดยต้องมีการตรวจสอบความซ้ำซ้อนของคำใน SenticNet ก่อนการจัดเก็บ



รูปที่ 2. กระบวนการสร้าง SenticNet

หากเป็นคำที่ยังไม่มีการจัดเก็บใน SenticNet จะต้องหาข้อสรุปของค่าความรู้สึกและค่าชั่วคราวอารมณ์โดยการนำค่าเหล่านั้นมาตรวจสอบกับอารมณ์ 6 ระดับ จาก 4 มิติ

ตามทฤษฎี Hourglass of Emotion โดยวิธีการวนซ้ำเพื่อหาคำนี้อยู่ในมิติใดในอารมณ์ 4 มิติ หลังจากนั้นนำมาผ่านกระบวนการ blending และ Spectral

association เพื่อหาค่าความรู้สึกและค่าชั่วคราวมณ โดย การ blending เป็นเทคนิคในการอนุมานค่าจากหลายๆ แหล่งข้อมูลพร้อมๆกัน ซึ่งเป็นการใช้ประโยชน์จาก แหล่งข้อมูลที่มีการซ้อนทับกันอยู่ใน AffectiveSpace เพื่อให้เกิดข้อมูลที่สามารถใช้งานร่วมกันได้ ส่วน Spectral association จะเป็นการหาค่าความรู้สึก ซึ่ง ผลลัพธ์ที่ได้หลังจากการทำ blending และ Spectral association คือ ค่าความรู้สึกและค่าชั่วคราวมณ กระบวนการดังกล่าว แสดงในรูปที่ 2

จากรูป จะเห็นได้ว่า กระบวนการหลักของ SenticNet จะมาจาก ส่วนแรก คือ SenticNet1 (สีขา) ซึ่งนำคำศัพท์ที่รวบรวมได้มาเข้ากระบวนการ Aligning SVD และลดมิติของคำ จนอยู่ในรูปของ AffectiveSpace แล้วเลือกคำที่ยังไม่ถูกจัดเก็บใน SenticNet มาหาค่าชั่วคราวมณ โดยการหามิติของอารมณ์จากคำเหล่านั้น แล้ว เข้าสู่กระบวนการ blending และ Spectral association เพื่อหาค่าชั่วคราวมณ แล้วนำมาผ่านกระบวนการ

Normalization ก่อนการจัดเก็บใน SenticNet ส่วนที่ 2 คือ SenticNet 2 (สีฟ้า) เป็นส่วนที่เพิ่มเติมจาก กระบวนการหลัก คือ ค่าความรู้สึกซึ่งจะเกิดขึ้นพร้อมๆกับการค่าชั่วคราวมณจาก SenticNet 1 และคำที่มีความหมายและอารมณ์เดียวกันโดยใช้หลักการแยก ความหมายจากข้อความ แล้วใช้ CF-IOF ในการหา ความหมายและระบุความหมายให้กับคำพร้อมทั้งจัดเก็บ ใน SenticNet ส่วนที่ 3 คือ SenticNet 3 (สีส้ม) จะใช้ คำศัพท์จาก COGBASE นำมาเข้ากระบวนการเดียวกับ SenticNet 1 ที่แสดงค่าความรู้สึกและค่าชั่วคราวมณ ส่วน ที่ 4 คือ SenticNet 4 (สีม่วง) เป็น SenticNet เวอร์ชัน ล่าสุดที่มีการใช้งานในปัจจุบัน โดยมีการเพิ่มเติมคำใหม่ที่มีความหมาย ที่เกิดจากการรวม verb concept และ noun concept และนำมาเข้ากระบวนการเดียวกับ SenticNet 1 และ SenticNet 2 เพื่อแสดงค่าความรู้สึก ค่าชั่วคราวมณ และคำที่มีความหมาย [2][6][7][8]

```
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:sentic="http://www.sentic.net/api/en/concept/"
  xmlns:float="http://www.w3.org/2001/XMLSchema#float">
  <rdf:Description rdf:about="http://www.sentic.net/api/en/concept/meet_friend">
    <rdf:type rdf:resource="http://www.sentic.net/api/en/concept/" />
    <text xmlns="http://www.sentic.net/">meet_friend</text>
    <semantics xmlns="http://www.sentic.net/">
      <concept xmlns="http://www.sentic.net/" rdf:resource="http://www.sentic.net/api/en/concept/meet_person"/>
      <concept xmlns="http://www.sentic.net/" rdf:resource="http://www.sentic.net/api/en/concept/chit_chat"/>
      <concept xmlns="http://www.sentic.net/" rdf:resource="http://www.sentic.net/api/en/concept/make_friend"/>
      <concept xmlns="http://www.sentic.net/" rdf:resource="http://www.sentic.net/api/en/concept/meet_girl"/>
      <concept xmlns="http://www.sentic.net/" rdf:resource="http://www.sentic.net/api/en/concept/socialize"/>
    </semantics>
    <sentic xmlns="http://www.sentic.net/">
      <pleasantness xmlns="http://www.sentic.net/" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.048</pleasantness>
      <attention xmlns="http://www.sentic.net/" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">-0.08</attention>
      <sensitivity xmlns="http://www.sentic.net/" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.036</sensitivity>
      <aptitude xmlns="http://www.sentic.net/" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0</aptitude>
    </sentic>
    <moodtags xmlns="http://www.sentic.net/">
      <concept xmlns="http://www.sentic.net/" rdf:resource="http://www.sentic.net/api/en/concept/joy"/>
      <concept xmlns="http://www.sentic.net/" rdf:resource="http://www.sentic.net/api/en/concept/surprise"/>
    </moodtags>
    <polarity xmlns="http://www.sentic.net/">
      <value xmlns="http://www.sentic.net/">positive</value>
      <intensity xmlns="http://www.sentic.net/" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.031</intensity>
    </polarity>
  </rdf:Description>
</rdf:RDF>
```

รูปที่ 3.แสดงโครงสร้างของ SenticNet ในรูปแบบ XML[5]

ในปัจจุบัน Senticnet ยังถูกสร้างเป็นไฟล์ RDF ในรูปแบบ XML ซึ่งแสดงในรูปที่ 3 และรูปแบบ API โดยใช้

ฐานข้อมูล SenticNet ที่ได้ นำมาสร้างเป็นแอปพลิเคชัน เพื่อให้ผู้ใช้สามารถดาวน์โหลดไปติดตั้งหรือเชื่อมต่อกับ

เครื่องมือต่างๆ ที่ใช้ในการแยกข้อความเป็นคำ แล้วประมวลผลค่าความรู้สึกด้วย SenticNet

จากรูปที่ 3 เป็น Sentic API ที่แสดงโครงสร้างเป็น XML จากตัวอย่าง เมื่อต้องการหาค่าความรู้สึกของคำว่า “meet friend” โดยใช้ Sentic API จะได้ชุดของข้อมูล 4 ชุด ดังนี้

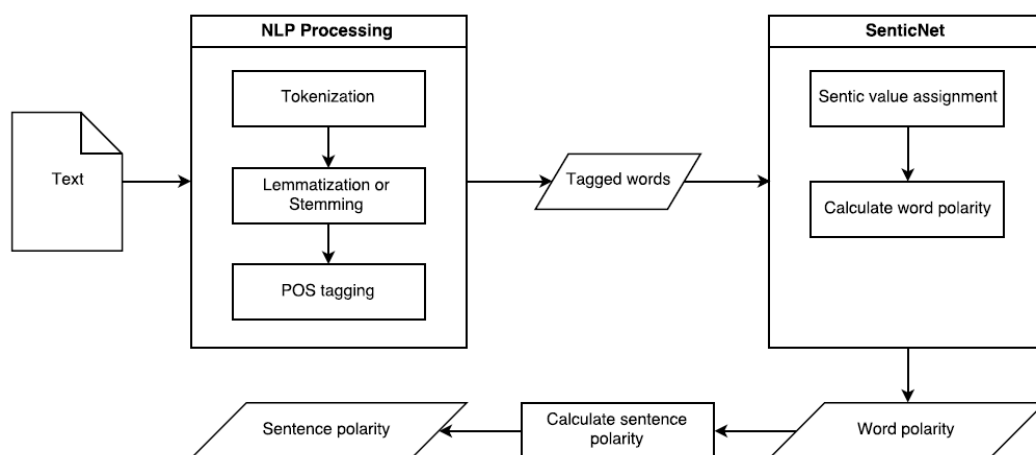
ชุดที่ 1 semantics คือ คำที่มีความหมายและอารมณ์เดียวกับคำที่ต้องการ ในตัวอย่างจะพบว่ามี 4 คำ คือ “meet person”, “chit chat”, “make friend”, “meet girl” และ “socialize”

ชุดที่ 2 sentics คือ ค่าความรู้สึก ของอารมณ์ 4 มิติ ในตัวอย่างจะพบว่า pleasantness มีค่าเท่ากับ 0.048 attention มีค่าเท่ากับ -0.08 sensitivity มีค่าเท่ากับ 0.036 และ aptitude มีค่าเท่ากับ 0

ชุดที่ 3 moodtags คือการบอกว่าการนี้จัดอยู่ในอารมณ์ระดับไหนจาก 6 ระดับ ในตัวอย่างจะพบว่าจะจัดอยู่ในอารมณ์ “joy” และ “surprise”

ชุดที่ 4 polarity คือ ขั้วอารมณ์ที่เป็นเชิงบวกหรือลบและค่าขั้วอารมณ์ ในตัวอย่างจะพบว่าแสดงขั้วอารมณ์ที่เป็นเชิงบวกและค่าขั้วอารมณ์เท่ากับ 0.031

#### 4. การนำ SenticNet ไปใช้งาน



รูปที่ 4. แสดงการนำ SenticNet ไปใช้งาน

เมื่อต้องการวิเคราะห์ความรู้สึกด้วย SenticNet จะต้องนำข้อความที่ต้องการวิเคราะห์ มาผ่านกระบวนการ NLP (Natural Language Processing) โดยต้องทำ Tokenization เพื่อแยกข้อความออกเป็นคำ ซึ่งในภาษาไทยต้องมีการตัดคำ (Word Segmentation) หลังจากนั้นนำมาทำ Lemmatization หรือ Stemming ซึ่งโดยปกติ กระบวนการ NLP จะเลือกใช้ Stemming เพราะสามารถทำงานได้รวดเร็วกว่า จากนั้นนำมากำกับชนิดของคำจากข้อความ (Part Of Speech Tagging)

อาทิเช่น คำนาม คำกริยา เป็นต้น หลังจากนั้นนำคำที่ผ่านกระบวนการ NLP แล้วมาหาขั้วอารมณ์จาก Senticnet ซึ่งสามารถดาวน์โหลดได้ฟรีทั้งแบบสแตนด์อโลนหรือแบบออนไลน์ผ่านทาง API โดย SenticNet จะค้นหาค่าความรู้สึกและค่าขั้วอารมณ์ โดยค่าขั้วอารมณ์ในระดับคำสามารถคำนวณได้[6] ดังนี้

$$p_c = \frac{Plsn(c) + |Attn(c)| - |Snst(c)| + Aptt(c)}{9} \quad (1)$$



แต่หากต้องการหาข้ออารมณ์ในระดับประโยค จะต้องนำค่าความรู้สึกของแต่ละคำที่ได้ มาคำนวณข้อความรู้สึกในระดับประโยค ดังนี้

$$p = \sum_{i=1}^N \frac{Plsn(ci) + |Attn(ci)| - |Snst(ci)| + Aptt(ci)}{9N} \quad (2)$$

เมื่อได้ค่าข้ออารมณ์แล้ว สามารถนำไปใช้งานในด้านต่างๆ ได้ต่อไป ซึ่งในปัจจุบันได้มีการนำ Senticnet ไปใช้งานในด้านต่างๆ ดังนี้

#### 4.1 ด้านสาธารณสุข

สถานพยาบาลได้นำ SenticNet มาใช้ในการวิเคราะห์ความรู้สึกจากข้อคิดเห็นหรือข้อเสนอแนะแบบออนไลน์ของผู้ที่มาใช้บริการในด้านต่างๆ ของสถานพยาบาล อาทิ เช่น ด้านสุขภาพ ด้านการอำนวยความสะดวก เป็นต้น ซึ่งข้อคิดเห็นเหล่านั้นถูกจัดเก็บไว้ในรูปแบบข้อความที่เป็นภาษาธรรมชาติ ไม่มีโครงสร้าง และมีจำนวนมาก โดยค่าข้ออารมณ์ที่ได้จาก SenticNet สามารถนำมาใช้ประเมินผลได้ว่า บริการในด้านใดบ้างที่ผู้ใช้บริการพึงพอใจ และมีด้านใดบ้างที่สถานพยาบาลควรดำเนินการปรับปรุงให้ดียิ่งขึ้น นอกจากนี้ยังมีการนำ SenticNet มาใช้ในการจัดอันดับสถานพยาบาลในพื้นที่ใกล้เคียง ในการบริการด้านสุขภาพเพื่อสร้างความน่าเชื่อถือให้กับสถานพยาบาลอีกด้วย[6][7]

#### 4.2 ด้านการตลาด

##### 4.2.1 การวิเคราะห์ความพึงพอใจของลูกค้า

ธุรกิจนำความคิดเห็นเกี่ยวกับความพึงพอใจของลูกค้า ที่มีต่อสินค้าและบริการ มาวิเคราะห์ความรู้สึกโดยใช้ SenticNet แล้วนำความคิดเห็นเหล่านั้นมาใช้ในการพยากรณ์ทางการตลาดหรือวิเคราะห์ข้อมูลทางการตลาด เช่น CRM หรือการประเมินความพึงพอใจของลูกค้าที่มีต่อสินค้าและบริการ เพื่อสร้างกลยุทธ์ทางการตลาดให้

เหนือกว่าคู่แข่ง และเพื่อเป็นส่วนช่วยในการตัดสินใจลงทุนในธุรกิจ และปรับปรุงสินค้าและบริการให้ดียิ่งขึ้น[6]

##### 4.2.2 การวิเคราะห์บทภาพยนตร์

การวิเคราะห์บทภาพยนตร์ เป็นการนำบทภาพยนตร์ ที่เป็นข้อความมาวิเคราะห์ความรู้สึก เพื่อหาค่าข้ออารมณ์ โดยใช้ SenticNet ในการประเมินผลและจำแนกประเภทภาพยนตร์ให้เหมาะสมกับผู้รับชม[4]

นอกจากนี้ ยังสามารถใช้ SenticNet ในงานต่างๆ ที่เป็นการกำกับหรือแสดงอารมณ์[5] อาทิเช่น e-business e-commerce e-governance e-security e-health e-learning, e-tourism เป็นต้น

### 5. บทสรุป

SenticNet เป็นแหล่งทรัพยากรที่ใช้ในการวิเคราะห์ความรู้สึกในระดับคำและระดับประโยค ที่สร้างขึ้นโดยใช้แหล่งทรัพยากรจากหลายแหล่ง ที่มีการรวบรวมคำศัพท์ที่เกี่ยวข้องกับอารมณ์ ความรู้สึก มาผ่านกระบวนการต่างๆ เพื่อหาค่าความรู้สึก ค่าข้ออารมณ์ และค่าที่มีความหมายเดียวกัน และจัดเก็บในฐานข้อมูลของ SenticNet พร้อมทั้งสร้างเป็น RDF/XML และ API ให้สามารถดาวน์โหลดไปใช้งานในด้านต่างๆ เช่น ด้านสาธารณสุข ด้านการตลาด เป็นต้น ซึ่ง SenticNet จะช่วยในการหาค่าความรู้สึกจากข้อความนั้นๆ และสามารถนำค่าความรู้สึกที่ได้ไปใช้ประโยชน์ต่อไป

### เอกสารอ้างอิง

- [1] C. Havasi, J. Alonso and R. Speer, "ConceptNet 3: a flexible, multilingual semantic network for common sense knowledge," in recent advances in natural language processing, Philadelphia, PA: John Benjamins, Sep. 2007, pp. 27-29.
- [2] B. W. Schuller, E. Cambria, R. Bajpai and S.

- Poria, "SenticNet 4: A Semantic Resource for Sentiment Analysis Based on Conceptual Primitives," In COLING, Dec. 2016, pp. 2666-2677.
- [3] A. Hussain, C. Eckl, C. Havasi and E. Cambria, "AffectiveSpace: Blending common sense and affective knowledge to perform emotive reasoning," WOMSA at CAEPIA, Seville, 2009, pp. 32-41.
- [4] A. Hussain, A. Livingstone and E. Cambria, "The hourglass of emotions. Cognitive behavioural systems," 2012, pp. 144-157.
- [5] A. Bera, A. Massa, C. Gui, E. Cambria, L. Yap and P. Vij, (2017). SenticNet [Online]. Available: <http://business.sentic.net>
- [6] A. Hussain , C. Havasi , E. Cambria and R. Speer, "SenticNet: A Publicly Available Semantic Resource for Opinion Mining," In AAAI fall symposium: commonsense knowledge, Vol. 10, No. 0, Nov. 2010.
- [7] A. Hussain , C. Havasi and E. Cambria, "SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis," In FLAIRS conference , May. 2012, pp. 202-207.
- [8] D. Olsher, D. Rajagopal and E. Cambria, "SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis," In Twenty-eighth AAAI conference on artificial intelligence, June. 2014.