การคัดเลือกคุณลักษณะสำคัญของคำบ่งอารมณ์จากเครือข่ายสังคมออนไลน์

นายศุภกิจ เชื้อธนะภิญโญ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2556
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

EMOTION TERM FEATURE SELECTION FROM ONLINE SOCIAL NETWORK

Mr. Supakit Chuetanapinyo

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Computer Engineering
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2013

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การคัดเลือกคุณลักษณะสำคัญของคำบ่งอารมณ์จากเครือข่ายสังคมออนไลน์
โดย	นายศุภกิจ เชื้อธนะภิญโญ
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.โปรดปราน บุณยพุกกณะ
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	รองศาสตราจารย์ ดร.อติวงศ์ สุชาโต

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์

(ศาสตราจารย์ ดร.บัณฑิต เอื้ออาภรณ์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์ ดร.โปรดปราน บุณยพุกกณะ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

(รองศาสตราจารย์ ดร.อติวงศ์ สุชาโต)

.....กรรมการภายนอกมหาวิทยาลัย

(ดร.ชูชาติ หฤไชยะศักดิ์)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ศุภกิจ เชื้อธนะภิญโญ : การคัดเลือกคุณลักษณะสำคัญของคำบ่งอารมณ์จากเครือข่ายสังคมออนไลน์. (EMOTION TERM FEATURE SELECTION FROM ONLINE SOCIAL NETWORK) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร.โปรดปราน บุณยพุกกณะ, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: รศ. ดร.อติวงศ์ สุชาโต, 92 หน้า.

ตั้งแต่สื่อทางสังคมออนไลน์กลายเป็นส่วนหนึ่งของชีวิตประจำวันของคนส่วนใหญ่ ได้ส่งผลกระทบโดยตรงต่อเทคโนโลยีคอมพิวเตอร์หลายแขนง รวมถึงการจำแนกของอารมณ์จากข้อความ เพราะผู้ใช้มีแนวโน้มในการใช้ภาษาไทยในสื่อสังคมออนไลน์เปลี่ยนแปลงไปจากเดิมเป็นอย่างมาก ดังนั้นในงานวิทยานิพนธ์นี้ผู้วิจัยจึงเสนอการคัดเลือกคุณลักษณะสำคัญในการจำแนกอารมณ์ และ การให้น้ำหนักของคำบ่งอารมณ์ พร้อมทั้งเพิ่มวิธีการจับคู่น้ำหนักที่เจ้าของข้อความระบุ ด้วยวิธีการให้น้ำหนักของคำแบบความถี่ของคำกับเดลต้า-ไคสแควร์ จากการทดลองเปรียบเทียบกับวิธีไค-สแควร์ดั้งเดิม พบว่ากรรมวิธีเดลต้า-ไคสแควร์ที่เสนอนี้ ได้ผลความแม่นยำในการจำแนกที่ 88.79% ซึ่งดีกว่าวิธีไค-สแควร์ที่ให้ผลความแม่นยำที่ 80.83% จาก 3,390 ข้อความจากเครือข่ายทางสังคมออนไลน์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาควิชา	วิศวกรรมคอมพิวเตอร์	ลายมือชื่อนิสิต .....
สาขาวิชา	วิศวกรรมคอมพิวเตอร์	ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....
ปีการศึกษา	2556	ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม .....

5370500621 : MAJOR COMPUTER ENGINEERING
KEYWORDS: EMOTION / TERM SELECTION / SUPERVISED TERM WEIGHTING / HUMAN EMOTION RECOGNITION

SUPAKIT CHUETANAPINYO: EMOTION TERM FEATURE SELECTION FROM ONLINE SOCIAL NETWORK. ADVISOR: ASST. PROF. DR.PROADPRAN P. PUNYABUKKANA, CO-ADVISOR: ASSOC. PROF. DR.ATIWONG SUCHATO, 92 pp.

Since social media has become part of our daily lives, it has posted some direct effect on computer technologies. This includes the classification of emotions from text. Specifically in Thai language, the trend in writing in social media has changed enormously. Therefore, this thesis proposes a means to improve term selection and term weighting by adding “owner-reader” matching algorithm to the Term Frequency and Delta-Chi Square supervised term weighting method. In our experiment, we compared our approach with the original Chi-Square method and found that our approach outperformed the baseline at 88.79% to 80.83% when tested on 3,390 documents from social network.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Department:	Computer Engineering	Student's Signature .....
Field of Study:	Computer Engineering	Advisor's Signature .....
Academic Year:	2013	Co-Advisor's Signature .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ สำเร็จลุล่วงไปได้ด้วยดีด้วยความช่วยเหลือจากทุกคนจึงทำให้วิทยานิพนธ์ฉบับนี้เสร็จสิ้นอย่างสมบูรณ์ในระยะเวลาที่กำหนดได้ ข้าพเจ้าขอขอบพระคุณ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก ผู้ช่วยศาสตราจารย์ ดร.โปรดปราน บุณยพุกกณะ และ อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม รองศาสตราจารย์ ดร.อติวงศ์ สุชาโต ซึ่งได้จุดประกายแนวคิดในการเริ่มต้นทำวิจัย แนะนำแนวทาง และ ได้สละเวลาอันมีค่าเพื่อให้คำปรึกษาเมื่อข้าพเจ้าต้องการได้เสมอ ข้าพเจ้าขอขอบคุณ ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ และ ดร.ชูชาติ หฤไชยะศักดิ์ ที่ให้คำแนะนำแนวคิดอันมีคุณค่าและถูกนำมาใช้ในการทำวิทยานิพนธ์นี้

ข้าพเจ้าขอขอขอบคุณ เพื่อนร่วมงานในห้องปฏิบัติการระบบภาษาพูด ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ให้ข้อแนะนำที่ดีกับข้าพเจ้า ให้กำลังใจ และสร้างรอยยิ้มให้ข้าพเจ้าในขณะที่ข้าพเจ้าท้อแท้ได้เสมอ

ข้าพเจ้าขอขอขอบคุณเพื่อนร่วมงานทุกคน และ ผู้บังคับบัญชาในสายงานที่คอยให้กำลังใจทำแบบสอบถามในงานวิจัย มีส่วนช่วยให้วิทยานิพนธ์ฉบับนี้สมบูรณ์ขึ้นมาได้

ข้าพเจ้าขอขอบคุณ คุณพ่อ คุณแม่ และครอบครัวที่เข้าใจและสนับสนุนการทำวิจัยในครั้งมาโดยตลอด ทำให้การจัดทำวิทยานิพนธ์ในครั้งนี้ประสบความสำเร็จได้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

1. บทนำ 1

1.1. ความเป็นมาและความสำคัญของปัญหา 1

1.2. วัตถุประสงค์ของการวิจัย 2

1.3. ขอบเขตของการวิจัย 2

1.4. ลำดับขั้นตอนในการเสนอผลการวิจัย 2

1.5. ประโยชน์ที่จะได้รับ 3

1.6. ผลงานตีพิมพ์จากวิทยานิพนธ์ 3

1.7. ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์ 3

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 4

2.1. ทฤษฎีและแนวคิดที่เกี่ยวข้อง 4

2.1.1. การแบ่งประเภทอารมณ์ 4

2.1.2. การตัดคำภาษาไทย 4

2.1.3. การหาความถี่ของการเกิดร่วมกันระหว่างคำ กับ อารมณ์ของข้อความ 5

2.1.4. การคำนวณค่าเฉลี่ยคะแนนของคำแบบเดลต้า-ไคสแควร์ 6

2.1.5. การสร้างแบบจำลองปริภูมิเวกเตอร์ 7

2.1.6. การคำนวณค่าน้ำหนักของคำ 8

2.1.7. การสร้างโมเดลฝึกฝน 13

2.1.8. การจำแนกประเภทของข้อมูล 14

2.2. งานวิจัยที่เกี่ยวข้อง 16

3. ขั้นตอนการดำเนินงานวิจัย 18

3.1. เครื่องมือที่ใช้ในการวิจัย 18

3.2. ขั้นตอนการดำเนินงานวิจัย 19

3.2.1. ขั้นตอนการจัดเตรียมข้อความภาษาไทย 20

CHULALONGKORN UNIVERSITY

หน้า

3.2.2. ขั้นตอนการคัดเลือกข้อความ ตรวจจับคำสะกดผิด และ ตัดคำภาษาไทย 26
3.2.3. ขั้นตอนการคัดเลือก และ การให้น้ำหนักของคำภาษาไทย 30
3.2.4. ขั้นตอนการสร้างโมเดลฝึกฝน และ การจำแนกประเภทอารมณ์บนข้อความ 35
4. การทดลอง และอภิปรายผล 38
4.1. การทดสอบแบบไขว้ข้าม (Cross Validation) 38
4.2. การทดสอบความแม่นยำ (Accuracy) บนแต่ละวิธีการให้น้ำหนักของคำ 40
4.3. การทดลองความแม่นยำของวิธีการคัดเลือกของคำ 56
4.4. การทดลองเปรียบเทียบความแม่นยำบนชุดข้อความระหว่างเฟสบุ๊ค กับ พันทิป 57
5. สรุปผลการวิจัย และข้อเสนอแนะ 58
5.1. สรุปผลการวิจัย 58
5.2. ข้อเสนอแนะ 59
รายการอ้างอิง 61
ประวัติผู้เขียนวิทยานิพนธ์ 92

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1
บทนำ

1. บทนำ

1.1. ความเป็นมาและความสำคัญของปัญหา

เทคโนโลยีการสื่อสาร หรือ การโต้ตอบระหว่างคอมพิวเตอร์กับมนุษย์ ได้ถูกพัฒนาอย่างต่อเนื่องจากนักวิจัย และหน่วยงานวิจัยต่างๆ ในหลายปีที่ผ่านมา ซึ่งเทคโนโลยีนั้นแบ่งการทำงานออกเป็น 3 ส่วนหลัก คือ กระบวนการรู้จำข้อมูล, กระบวนการประมวลผลข้อมูล และ กระบวนการสังเคราะห์ข้อมูล เทคโนโลยีที่ยังเป็นที่ท้าทายสำหรับเหล่าบรรดานักวิจัย นั้นคือ เทคโนโลยีการรู้จำอารมณ์ของมนุษย์ (Human Emotion Recognition, HER) [1] เป็นเทคโนโลยีเพื่อให้ระบบได้เรียนรู้อารมณ์ของมนุษย์ ซึ่งการเรียนรู้นั้นมีด้วยกันหลายวิธีการ อาทิเช่น การเรียนรู้จากภาพ, การเรียนรู้จากเสียง และ การเรียนรู้จากข้อความ ในการเรียนรู้อารมณ์จากข้อความ ข้อมูลที่ระบบจะเรียนรู้ได้นั้นจะประกอบด้วยข้อความ กับ อารมณ์ของข้อความ บางครั้งจำเป็นต้องใช้ปริมาณข้อมูลที่มีจำนวนมาก เช่น ข้อความจากเครือข่ายสังคมออนไลน์ เพื่อให้ระบบได้เรียนรู้อย่างคลอบคลุมกับคำที่เกิดขึ้น อย่างไรก็ตามการระบุอารมณ์บนข้อความโดยมนุษย์ ในปริมาณข้อมูลที่มีจำนวนมาก ต้องใช้เวลาในการจัดเตรียมชุดข้อมูลสำหรับฝึกฝน ดังนั้นระบบการจำแนกอารมณ์บนข้อความที่เหมาะสม สามารถช่วยลดเวลาในขั้นตอนการระบุอารมณ์บนข้อความโดยมนุษย์ลงไปได้ ทำให้เกิดงานวิจัยด้านการจำแนกอารมณ์บนข้อความ จากเครือข่ายสังคมออนไลน์

ในงานวิจัยที่เกี่ยวข้องกับการจำแนกอารมณ์บนข้อความที่ผ่านมา วิธีการคัดเลือกคุณลักษณะสำคัญที่นิยมใช้กันอย่างแพร่หลาย คือ วิธีการกำจัดคำหยุด (Stop-Word Removal Method) [1] และ วิธีการคัดเลือกหน้าที่คำ (Part of Speech, POS) [2] [3] ปัญหาที่เกิดขึ้นจากวิธีนี้คือ หน้าที่ของคำบางส่วนที่ไม่ได้ถูกเลือกมาจะไม่ได้ทำหน้าที่เป็นคุณลักษณะสำคัญสำหรับ การจำแนกอารมณ์บนข้อความ ดังนั้นในงานวิจัยนี้จะคำนึงถึงทุกคุณลักษณะสำคัญที่เกิดขึ้น โดยการให้คะแนนทุกคุณลักษณะสำคัญแบบเดลต้า-ไคสแควร์ (Delta Chi-Square, δ-$\chi 2$) [4] แล้วทำการเรียงคะแนนจากมากไปน้อย เพื่อให้ได้ชุดคุณลักษณะสำคัญที่เหมาะสมสำหรับการจำแนกอารมณ์บนข้อความ

ส่วนในงานวิจัยที่เกี่ยวข้องกับการให้น้ำหนักของคำที่ผ่านมา เป็นการให้ค่าน้ำหนักกับคำที่มีคุณลักษณะเฉพาะกับอารมณ์ของข้อความ สามารถแบ่งเป็นสองวิธีหลัก [5] คือ วิธีการให้น้ำหนักของคำแบบไม่มีผู้สอน (Unsupervised Term Weighting Method) และ วิธีการให้น้ำหนักของคำแบบมีผู้สอน (Supervised Term Weighting Method) ในส่วนวิธีการให้น้ำหนักของคำแบบมีผู้สอน ส่วนใหญ่ได้รับแนวคิด [6] มาจากความถี่ของคำ กับ วิธีการคัดเลือกคุณลักษณะสำคัญ ดังนั้นในวิจัยนี้ จึงได้นำเสนอวิธีการให้น้ำหนักความถี่ของคำ กับ วิธีการคัดเลือกแบบเดลต้า-ไคสแควร์ เข้าไปด้วย

1.2. วัตถุประสงค์ของการวิจัย

วัตถุประสงค์ของการวิจัย คือ เสนอวิธีการคัดเลือกคุณลักษณะสำคัญ และ การให้น้ำหนักของคำบ่งอารมณ์ จากเครือข่ายสังคมออนไลน์ เพื่อเพิ่มประสิทธิภาพในการจำแนกอารมณ์บนข้อความ

1.3. ขอบเขตของการวิจัย

1.3.1. ข้อความที่คัดเลือกมาต้องเป็นข้อความภาษาไทยเท่านั้น

1.3.2. ข้อความคัดเลือกมาจากเครือข่ายสังคมออนไลน์ เช่น เฟสบุ๊ค [7], พันทิป [8]

1.3.3. ข้อความที่คัดเลือกมา เจ้าของข้อความสามารถระบุอารมณ์บนข้อความได้

1.3.4. อารมณ์บนข้อความ อ้างอิงจากพื้นฐานของอารมณ์ 6 ประเภท [9] ได้แก่ ร่าเริง, กลัว, ประหลาดใจ, เสียใจ, รังเกียจ และ โกรธ

1.3.5. เสนอวิธีการคัดเลือกคุณลักษณะสำคัญ และ การให้น้ำหนักของคำ ด้วยวิธีเดลต้า-ไคสแควร์ (Delta-Chi Square, δ-$\chi 2$) [4] และ มีการเปรียบเทียบกับวิธีการอื่นซึ่งอยู่ในขั้นตอนการดำเนินงานวิจัย

1.3.6. ทดสอบแบบไขว้ข้าม (K-Cross Validation) และ ความแม่นยำ (Accuracy) ของการจำแนกอารมณ์บนข้อความ ด้วยอัลกอริทึมนาอีฟเบย์ (Naïve Bayes) [10]

1.4. ลำดับขั้นตอนในการเสนอผลการวิจัย

<u>1.4.1 ขั้นตอนศึกษาเบื้องต้น</u>

1.4.1.1 ศึกษางานวิจัยที่เกี่ยวกับการคัดเลือกคุณลักษณะสำคัญของคำที่บ่งอารมณ์

1.4.1.2 ศึกษางานวิจัยที่เกี่ยวกับการจำแนกอารมณ์บนข้อความ

1.4.1.3 ศึกษาอัลกอริทึมในการจำแนกอารมณ์บนข้อความ

1.4.1.4 ศึกษาการแปลงข้อความแบบจำลองปริภูมิเวกเตอร์ (Vector Space Model, VSM) [1]

1.4.1.5 ศึกษาระบบการจัดเก็บและจัดเตรียมแบบสอบถามสำหรับระบุอารมณ์บนข้อความ

<u>1.4.2 ขั้นตอนการออกแบบระบบ</u>

1.4.2.1 ออกแบบการทำงานของระบบ

1.4.2.2 เลือกใช้เครื่องมือในการวิจัยที่เหมาะสม

1.4.2.3 ออกแบบการทดลอง

<u>1.4.3 ขั้นตอนการทดลอง</u>

1.4.3.1 ทดสอบการทำงานของระบบ

1.4.3.2 บันทึกผลการทดลอง

1.4.3.3 วิเคราะห์ผลการทดลอง

<u>1.4.4 สรุปผลและเรียบเรียงวิทยานิพนธ์</u>

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.5. ประโยชน์ที่จะได้รับ

1.5.1. เพิ่มความแม่นยำในการจำแนกอารมณ์บนข้อความ โดยการให้น้ำหนักของคำบนข้อความที่นำเสนอในงานวิจัยนี้ เทียบกับงานวิจัยที่ผ่านมา ซึ่งแสดงในตารางที่ 4.2, 4.7

1.5.2. เพิ่มประสิทธิภาพในการคัดเลือกคำ โดยการให้คะแนน และ คัดเลือกจากอันดับคะแนนคำ เทียบกับการคัดเลือกคำตามหน้าที่คำที่ใช้ในงานวิจัยที่ผ่านมา ซึ่งแสดงในตารางที่ 4.3-4.6

1.5.3. นำเสนอวิธีการการเรียนรู้แบบมีผู้สอนจากทั้งเจ้าของข้อความ และ ไม่ใช่เจ้าของข้อความ อีกทั้งงานวิจัยนี้ เพิ่มเติมวิธีการจับคู่ระหว่างเจ้าของข้อความ และ ไม่ใช่เจ้าของข้อความ ซึ่งแสดงในตารางที่ 4.8

1.5.4. วิเคราะห์ความแม่นยำการจำแนกอารมณ์บนข้อความ จากการปรับพารามิเตอร์จำนวนคำบนข้อมูลชุดฝึกฝน ซึ่งแสดงในตารางที่ 4.2

1.5.5. สามารถนำระบบนี้ไปประยุกต์ใช้กับการจำแนกอารมณ์บนข้อความ ในเครือข่ายสังคมออนไลน์ ซึ่งแสดงในตารางที่ 4.12

1.6. ผลงานตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้ตีพิมพ์เป็นบทความทางวิชาการในหัวข้อเรื่อง “Emotion Term Feature Selection” จัดทำโดย “Supakit Chuetanapinyo, Proadpran Punyabukkana, Atiwong Suchato” ถูกนำเสนอในงานประชุมวิชาการ “Regional Conference on Computer and Information Engineering: RCCIE'2013” ณ โรงแรมสุโกศล จังหวัดกรุงเทพมหานคร ประเทศไทย ในวันที่ 23 สิงหาคม 2556

1.7. ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์

ในวิทยานิพนธ์ฉบับนี้มีการจัดเรียงเนื้อหาแบ่งออกเป็น 5 บท ตามลำดับดังนี้ บทที่ 1 บทนำ กล่าวถึงความเป็นมาและความสำคัญของปัญหา, วัตถุประสงค์ของการวิจัย, ขอบเขตของการวิจัย, ลำดับขั้นตอนในการเสนอผลการวิจัย และ ผลงานตีพิมพ์จากวิทยานิพนธ์ บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง กล่าวถึงทฤษฎีและแนวคิดที่เกี่ยวข้อง ประกอบด้วย การแบ่งประเภทอารมณ์บนข้อความ, การตัดคำภาษาไทย, การหาความถี่ของการเกิดร่วมกันระหว่างคำ กับ อารมณ์ของข้อความ, การคำนวณค่าคะแนนเฉลี่ยของคำแบบไคสแควร์, การสร้างแบบจำลองปริภูมิเวกเตอร์, การคำนวณค่าน้ำหนักของคำ, การสร้างโมเดลฝึกฝน และ การจำแนกประเภทของข้อมูล อีกส่วนคือ งานวิจัยที่เกี่ยวข้อง บทที่ 3 ขั้นตอนการดำเนินงานวิจัย กล่าวถึง เครื่องมือที่ใช้ในงานวิจัย ขั้นตอนการดำเนินงานวิจัย ประกอบด้วยขั้นตอนการจัดเตรียมข้อความภาษาไทย, ขั้นตอนการคัดเลือกข้อความ, ขั้นตอนการคัดเลือกคำ กับการให้น้ำหนักของคำ และ ขั้นตอนการสร้างโมเดลสำหรับฝึกฝนกับการจำแนกประเภทอารมณ์บนข้อความ บทที่ 4 การทดลอง และอภิปรายผล กล่าวถึง การทดลองของวิทยานิพนธ์นี้ และผลการทดลอง บทที่ 5 บทสรุปผลการวิจัยและข้อเสนอแนะ กล่าวถึงการสรุปผลการวิจัย ข้อเสนอในจุดเด่นจุดด้อยของงานวิจัย และ งานวิจัยในอนาคต

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1. ทฤษฎีและแนวคิดที่เกี่ยวข้อง

งานวิจัยนี้ทฤษฎีที่เกี่ยวข้องแบ่งออกเป็น การแบ่งประเภทอารมณ์, การตัดคำภาษาไทย, การหาความถี่ของการเกิดร่วมกันระหว่างคำ กับ อารมณ์ของข้อความ, การคำนวณค่าคะแนนเฉลี่ยของคำแบบไคส-แควร์, การสร้างแบบจำลองปริภูมิเวกเตอร์, การคำนวณค่าน้ำหนักของคำ การสร้างโมเดลฝึกฝน และ การจำแนกประเภทของข้อมูล

2.1.1. การแบ่งประเภทอารมณ์

ในหลายงานวิจัยที่ผ่านมา มีความพยายามในการหาตัววัดที่บ่งบอกอารมณ์ของมนุษย์ที่เกิดขึ้นในหลายวิธีการ [19] และ หนึ่งในนั้นคือ การแสดงออกทางรูปหน้าของมนุษย์ (Facial Expressions) [20] ในงานวิจัยนี้ การแบ่งประเภทอารมณ์ได้อ้างอิงจาก Paul Ekman [9] ซึ่งจะหาความสัมพันธ์ระหว่างการแสดงออกทางรูปหน้ากับการแสดงอารมณ์ที่เกิดขึ้นของมนุษย์ [21] ซึ่งสามารถแบ่งออกเป็นหกอารมณ์ด้วยกัน ได้แก่ ร่าเริง, กลัว , ประหลาดใจ, เสียใจ, รังเกียจ และ โกรธ

2.1.2. การตัดคำภาษาไทย

เนื่องจากลักษณะข้อความในภาษาไทย มีลักษณะการเขียนด้วยตัวอักษรหรือคำที่ติดกันอย่างต่อเนื่อง [11] ดังนั้นการตัดคำภาษาไทยจึงเป็นตัวช่วยให้ระบบสามารถวิเคราะห์คำที่อยู่บนข้อความนั้นได้ ในงานวิจัยนี้ได้เลือกใช้โปรแกรม TLex (Thai Lexeme Analyser) [12] โดยใช้อัลกอริทึมคอนดิชั่นนอลแรนดอมฟิลด์ (Conditional Random Fields) [22] ในการฝึกฝนชุดข้อมูลบนฐานข้อมูล InterBEST 2009 [12] และ คุณลักษณะ (Feature) ที่ใช้ในการตัดคำ คือ การผสมผสานของคุณลักษณะ (Combined Feature) [23] ซึ่งประกอบด้วย ตัวอักษร (Char) กับ ชนิดของตัวอักษร (Char-Type) ในงานวิจัยที่ผ่าน [12] ให้ผลความแม่นยำอยู่ที่ 93.63% ส่วนในงานวิจัยนี้ ให้ผลความแม่นยำอยู่ที่ 96.53% ดังแสดงตาราง 2.1-2.2

ตารางที่ 2.1 ตารางแสดงจำนวนการตัดคำที่ถูกต้อง, ไม่ถูกต้อง และจำนวนการตัดคำทั้งหมด

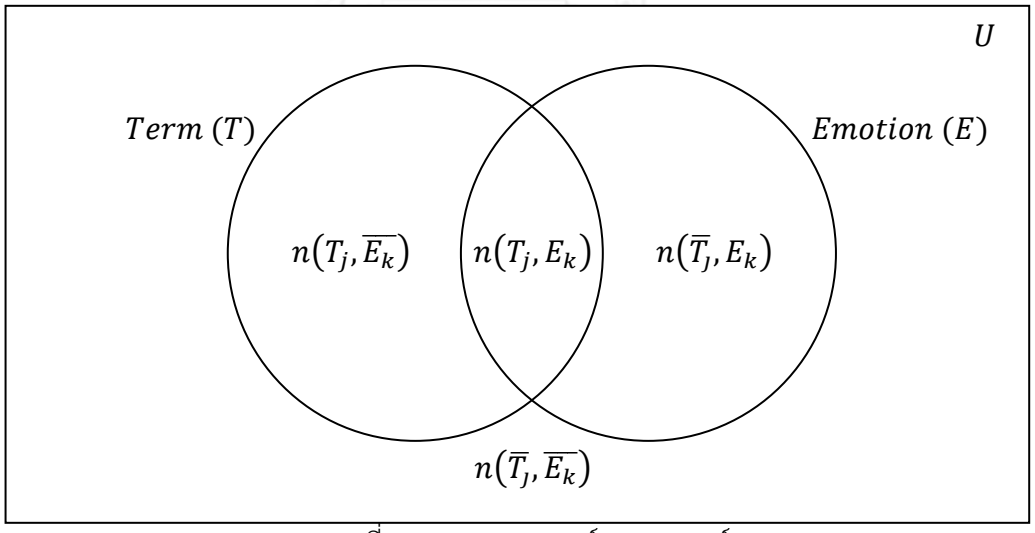
	จำนวนครั้ง	คิดเป็นเปอร์เซ็นต์
จำนวนการตัดคำ ทั้งหมด	26,828	100.00 %
จำนวนการตัดคำ ที่ถูกต้อง	25,898	96.53 %
จำนวนการตัดคำ ที่ไม่ถูกต้อง	930	3.47 %

ตารางที่ 2.2 ตารางแสดงจำนวนการตัดคำที่ไม่ถูกต้อง และ สาเหตุที่ตัดคำไม่ถูกต้อง

สาเหตุที่ตัดคำไม่ถูกต้อง	จำนวนครั้ง	คิดเป็น %	ตัวอย่าง
ตัดคำผิด	114	0.425 %	ที่|ผ่าน|มา| |50| |ปี| |ไทย|**นักการ|เมือง|**โกงกิน|
มีชื่อเฉพาะเกิดขึ้น	57	0.212 %	แล้ว|ผู้|หญิง|ชอบ|**กัน|ดั้ม|**จะ|แปลก|ตรง|ไหน|
ตัดคำไม่สมบูรณ์	267	0.995 %	อ่าน|แล้ว|ไอ|คู่|หมั้น|นี่|น่า|**โดนยำทีนจิง|**
มีอีโมติคอน	492	1.834 %	ขอบคุณ|ครับ|เตรียม|ได้|ซื้อ|แล้ว|สิ|นี่| |^| |^|

2.1.3. การหาความถี่ของการเกิดร่วมกันระหว่างคำ กับ อารมณ์ของข้อความ

การหาความถี่ของการเกิดร่วมกันระหว่างคำ กับ อารมณ์ของข้อความ [13] เป็นการหาความสัมพันธ์ระหว่างคำ กับ อารมณ์ของข้อความ โดยการนับความถี่ในการเกิดขึ้นร่วมกันในข้อความนั้น ซึ่งสามารถแสดงความสัมพันธ์ด้วยแผนภาพเวนน์-ออยเลอร์ [24] ดังภาพที่ 2.1



ภาพที่ 2.1 แผนภาพเวนน์ - ออยเลอร์
แสดงความสัมพันธ์ระหว่างคำ กับ อารมณ์บนข้อความ

$n(T_j, E_k)$ คือ จำนวนของข้อความมีคำที่ j และ เป็นอารมณ์ที่ k
$n(T_j, \overline{E_k})$ คือ จำนวนของข้อความมีคำที่ j และ ไม่เป็นอารมณ์ที่ k
$n(\overline{T_j}, E_k)$ คือ จำนวนของข้อความไม่มีคำที่ j และ เป็นอารมณ์ที่ k
$n(\overline{T_j}, \overline{E_k})$ คือ จำนวนของข้อความไม่มีคำที่ j และ ไม่เป็นอารมณ์ที่ k

2.1.4. การคำนวณค่าเฉลี่ยคะแนนของคำแบบเดลต้า-ไคสแควร์

การหาค่าคะแนนของคำ คือ กระบวนการหาค่าความดี หรือ ความสามารถในการแยกแยะประเภทอารมณ์บนข้อความ ซึ่งถ้าค่าคะแนนของคำมีค่าเป็นศูนย์ หมายถึง คำนั้นไม่มีความสามารถในการแยกแยะประเภทอารมณ์ แต่ถ้าค่าคะแนนของคำมีค่าไม่เป็นศูนย์ หมายถึง คำนั้นจะมีความสามารถในการแยกแยะประเภทอารมณ์ ตามค่าคะแนนของคำ เนื่องจากในงานวิจัยนี้ คำที่ปรากฏขึ้นบางคำมีลักษณะการปรากฏขึ้นในทุกอารมณ์บนข้อความ ดังนั้นในงานวิจัยนี้ จึงได้นำเสนอวิธีการให้คะแนนของคำแบบเดลต้า-ไคสแควร์ ซึ่งสามารถช่วยแก้ปัญหาบางคำที่มีลักษณะการปรากฏขึ้นในทุกอารมณ์บนข้อความได้ อีกทั้งยังนำแนวคิดดังกล่าวมาประยุกต์ใช้กับการให้น้ำหนักของคำ เพื่อให้อัลกอริทึมมีบทบาทในการคัดเลือกการให้น้ำหนักของคำบนข้อความ โดยผ่านพารามิเตอร์เดลต้า

การคำนวณค่าคะแนนของคำแบบเดลต้า-ไคสแควร์ [4] ประกอบด้วย 2 ส่วน คือ การคำนวณไคสแควร์ กับ การคำนวณเดลต้า 1) การคำนวณไคสแควร์ เป็นการคำนวณโดยการวัดความบกพร่องความอิสระต่อกัน (The lack of independence) [13] ระหว่างคำกับ อารมณ์ของข้อความ ดังตามสมการที่ 1 และ 2) การคำนวณเดลต้า เป็นการคำนวณโดยการวัดคุณลักษณะเฉพาะของคำที่มีต่ออารมณ์ของข้อความ ด้วยการวัดค่าความแตกต่างระหว่าง $n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k})$ กับ $n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k)$ [4] ดังตามสมการที่ 2 และ 3

$$Score_{\chi 2}(T_j, E_k) = \frac{N \times \left[n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k}) - n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k)\right]^2}{n(T_j) \times n(\overline{T_j}) \times n(E_k) \times n(\overline{E_k})} \quad (1)$$

$$Score_{\delta-\chi 2}(T_j, E_k) = \frac{\delta \times N \times \left[n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k}) - n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k)\right]^2}{n(T_j) \times n(\overline{T_j}) \times n(E_k) \times n(\overline{E_k})} \quad (2)$$

$$\delta = \begin{cases} 1, & n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k}) - n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k) > 0 \\ 0, & n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k}) - n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k) \leq 0 \end{cases} \quad (3)$$

โดยที่ $Score(T_j, E_k)$ คือ ค่าคะแนนของคำที่ j กับ อารมณ์ที่ k
δ คือ ค่าเดลต้า มีค่าเท่ากับ 1 หรือ 0
N คือ จำนวนข้อความบนข้อมูลชุดฝึกฝน

$n(T_j)$, $n(\overline{T_j})$ คือ จำนวนของข้อความมีคำที่ j และ ไม่มีคำที่ j
$n(E_k)$, $n(\overline{E_k})$ คือ จำนวนของข้อความเป็นอารมณ์ที่ k และ ไม่เป็นอารมณ์ที่ k

$n(T_j, E_k)$ คือ จำนวนของข้อความมีคำที่ j และ เป็นอารมณ์ที่ k
$n(T_j, \overline{E_k})$ คือ จำนวนของข้อความมีคำที่ j และ ไม่เป็นอารมณ์ที่ k
$n(\overline{T_j}, E_k)$ คือ จำนวนของข้อความไม่มีคำที่ j และ เป็นอารมณ์ที่ k
$n(\overline{T_j}, \overline{E_k})$ คือ จำนวนของข้อความไม่มีคำที่ j และ ไม่เป็นอารมณ์ที่ k

จากการคำนวณค่า $Score_{\delta-\chi2}(T_j, E_k)$ แต่ละคำ T_j จะได้ค่า $Score_{\delta-\chi2}(T_j, E_k)$ จำนวน 6 ค่า ตามประเภทอารมณ์ของข้อความ การหาคะแนนของคำในงานวิจัยนี้จะใช้การหาค่าเฉลี่ยแบบมีน้ำหนักของค่า $Score_{\delta-\chi2}(T_j, E_k)$ [13] ดังตามสมการที่ 4

$$Score_{average}(T_j) = \sum_{k=1}^{K} \{P(E_k) \times Score_{\delta-\chi2}(T_j, E_k)\} \quad (4)$$

โดยที่ $Score_{average}(T_j)$ คือ ค่าคะแนนเฉลี่ยของคำที่ j
$Score_{\delta-\chi2}(T_j, E_k)$ คือ ค่าคะแนนของคำที่ j กับ อารมณ์ที่ k
$P(E_k)$ คือ ความน่าจะเป็นข้อความจะเป็นอารมณ์ที่ k
K คือ จำนวนประเภทอารมณ์ บนข้อความ

2.1.5. การสร้างแบบจำลองปริภูมิเวกเตอร์

การสร้างแบบจำลองปริภูมิเวกเตอร์ คือ กระบวนการแปลงข้อความภาษาไทยเป็นแบบจำลองปริภูมิเวกเตอร์ [25] เริ่มจากการสร้างเซตของคำ แต่ละคุณลักษณะคำ (t_j) แทนคุณลักษณะคำบนข้อมูลชุดฝึกฝน จำนวน M คำ ดังตามสมการที่ 5

$$T = \{T_1, T_2, T_3, \dots, T_M\} \quad (5)$$

โดยที่ T คือ เซตของคำบนข้อมูลชุดฝึกฝน
T_j คือ คำลำดับที่ j
M คือ จำนวนคำ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากนั้นทำการสร้างเซตของข้อความ (D) แต่ละข้อความ (d_i) แทนแต่ละข้อความบนข้อมูลชุดฝึกฝน จำนวน N ข้อความ ดังตามสมการที่ 6

$$D = \{d_1, d_2, d_3, \dots, d_N\} \quad (6)$$

โดยที่ D คือ เซตของข้อความบนข้อมูลชุดฝึกฝน
d_i คือ ข้อความลำดับที่ i
N คือ จำนวนข้อความ

จากเซตของข้อความ (D) แต่ละข้อความ (d_i) จะถูกแปลงเป็นเวกเตอร์ ซึ่งประกอบด้วยค่าน้ำหนักของคำ (w_{ij}) หรือ เรียกว่า “Bag-of-Words” [26] ดังตามสมการที่ 7

$$\vec{d}_i = \langle w_{i1}, w_{i2}, w_{i3}, \dots, w_{ij} \rangle \quad (7)$$

โดยที่ $\vec{d}_i$ คือ เวกเตอร์ของข้อความที่ i
w_{ij} คือ ค่าน้ำหนักของคำที่ j บนข้อความที่ i

จากเวกเตอร์ของข้อความ $(\vec{d}_i)$ ซึ่งประกอบด้วยค่าน้ำหนักของคำ (w_{ij}) จะสามารถสร้างแบบจำลองปริภูมิเวกเตอร์ [25] ได้ดังตามสมการที่ 8

$$\begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \dots \\ \\ d_i \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & \dots & w_{1j} \\ w_{21} & w_{22} & w_{23} & \dots & w_{2j} \\ w_{31} & w_{32} & w_{33} & \dots & w_{31} \\ \dots & \dots & \dots & & \dots \\ & & & & \\ w_{i1} & w_{i2} & w_{i1} & \dots & w_{ij} \end{bmatrix} \quad (8)$$

โดยที่ d_i คือ ข้อความลำดับที่ i
w_{ij} คือ ค่าน้ำหนักของคำที่ j บนข้อความที่ i

2.1.6. การคำนวณค่าน้ำหนักของคำ

การคำนวณค่าน้ำหนักของคำ คือ กระบวนการให้ค่าน้ำหนักกับคำที่มีคุณลักษณะเฉพาะกับอารมณ์ หรือ มีความสามารถในการแยกแยะประเภทอารมณ์บนข้อความ [5] การให้ค่าน้ำหนักของคำจะถูกพิจารณาออกเป็น 3 ส่วน คือ ส่วนความถี่ของคำ (Term Frequency, tf) [27], ส่วนการคำนวณแบบเดลต้า-ไคสแควร์ (Delta-Chi Square, δ-$\chi 2$) [4] และ ส่วนการคำนวณค่าน้ำหนักให้อยู่บนบรรทัดฐานเดียวกัน (Normalization) [5]

ส่วนความถี่ของคำ (Term Frequency, tf) [27] คือ การนับความถี่ของคำที่ปรากฏในแต่ละข้อความ ดังตามสมการที่ 9

$$tf = n \tag{9}$$

โดยที่ n คือ ความถี่ของคำที่ปรากฏในแต่ละข้อความ

ส่วนการคำนวณแบบเดลต้า-ไคสแควร์ (Delta-Chi Square, δ-$\chi 2$) [4] คือ การวัดความบกพร่องความอิสระต่อกัน (The lack of independence) ระหว่างคำ กับ อารมณ์ของข้อความ และ การวัดคุณลักษณะเฉพาะของคำที่มีต่ออารมณ์ของข้อความ ดังตามสมการที่ 10 และ 11

$$W_{\delta-\chi 2}(T_j, E_k) = \frac{\delta \times N \times \left[n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k}) - n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k)\right]^2}{n(T_j) \times n(\overline{T_j}) \times n(E_k) \times n(\overline{E_k})} \tag{10}$$

$$\delta = \begin{cases} 1, & n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k}) - n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k) > 0 \\ 0, & n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k}) - n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k) \leq 0 \end{cases} \tag{11}$$

โดยที่ $W_{tf.\delta-\chi 2}$ คือ ค่าน้ำหนักของคำ ที่ถูกคำนวณแบบเดลต้า-ไคสแควร์
δ คือ ค่าเดลต้า มีค่าเท่ากับ 1 หรือ 0
N คือ จำนวนข้อความบนข้อมูลชุดฝึกฝน
$n(T_j), n(\overline{T_j})$ คือ จำนวนของข้อความมีคำที่ j และ ไม่มีคำที่ j
$n(E_k), n(\overline{E_k})$ คือ จำนวนของข้อความเป็นอารมณ์ที่ k และไม่เป็นอารมณ์ที่ k
$n(T_j, E_k)$ คือ จำนวนของข้อความมีคำที่ j และ เป็นอารมณ์ที่ k
$n(T_j, \overline{E_k})$ คือ จำนวนของข้อความมีคำที่ j และ ไม่เป็นอารมณ์ที่ k
$n(\overline{T_j}, E_k)$ คือ จำนวนของข้อความไม่มีคำที่ j และ เป็นอารมณ์ที่ k
$n(\overline{T_j}, \overline{E_k})$ คือ จำนวนของข้อความไม่มีคำที่ j และ ไม่เป็นอารมณ์ที่ k

จากส่วนความถี่ของคำ กับ ส่วนการคำนวณแบบเดลต้า-ไคสแควร์ จะสามารถสร้างสมการการให้น้ำหนักของคำด้วยวิธีการคำนวณความถี่ของคำ กับ เดลต้า-ไคสแควร์ (Delta-Chi Square, δ-$\chi 2$) ดังตามสมการที่ 12

$$W_{tf.\delta-\chi 2} = tf \times \frac{\delta \times N \times \left[n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k}) - n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k)\right]^2}{n(T_j) \times n(\overline{T_j}) \times n(E_k) \times n(\overline{E_k})} \tag{12}$$

ส่วนการคำนวณค่าน้ำหนักให้อยู่บนบรรทัดฐานเดียวกัน (Normalization) คือ การปรับค่าน้ำหนักให้มีความเท่าเทียมกัน [5] บนแต่ละข้อความ ในงานวิจัยนี้ได้เลือกใช้วิธี Cosine Normalization ดังตามสมการที่ 13

$$W_{Normalized}(d_i, t_j) = \frac{W(d_i, t_j)}{\sqrt{\sum_{j=1}^{m}[W^2(d_i, t_j)]}} \quad (13)$$

โดยที่ $W_{Normalized}$ คือ ค่าน้ำหนักที่ถูกคำนวณอยู่บนบรรทัดฐานเดียวกัน
W คือ ค่าน้ำหนัก
d_i, t_j คือ ข้อความที่ i กับ คำที่ j

ในส่วนต่อไป จะแสดงตัวอย่างการคำนวณค่าน้ำหนักของคำ แบบความถี่ของคำ กับ เดลต้า-ไคสแควร์ ที่นำเสนอในงานวิจัยนี้

ตัวอย่างข้อความ:
“อยาก | ดู | แต่ | ดู | ไม่ | ได้ | เพราะ | พรุ่ง | นี้ | ต้อง | ไป | โรง | เรียน |”

อารมณ์บนข้อความ (จากการโหวตของอาสาสมัคร):
“เสียใจ”

การคำนวณค่าน้ำหนักของคำ แบบความถี่ของคำ กับ เดลต้า-ไคสแควร์จะแบ่งการคำนวณออกเป็นสามขั้นตอนหลัก คือ 1) ค่าความถี่ของคำ 2) ค่าของไคสแควร์ และ 3) ค่าของเดลต้า

ส่วนแรก ค่าความถี่ของคำ จะทำการนับจำนวนของคำนั้น ที่ปรากฏบนหนึ่งข้อความดังสมการ 9 จากตัวอย่างข้อความจะได้ว่า

$tf(\text{อยาก}) = 1,$ $tf(\text{ดู}) = 2,$ $tf(\text{แต่}) = 1,$ $tf(\text{ไม่}) = 1,$
$tf(\text{ได้}) = 1,$ $tf(\text{เพราะ}) = 1,$ $tf(\text{พรุ่ง}) = 1,$ $tf(\text{นี้}) = 1,$
$tf(\text{ต้อง}) = 1,$ $tf(\text{ไป}) = 1,$ $tf(\text{โรง}) = 1,$ $tf(\text{เรียน}) = 1$

ส่วนที่สอง ค่าของไคสแควร์ จะทำการวัดความบกพร่องความอิสระต่อกัน (The lack of independence) ระหว่างคำ กับ อารมณ์ของข้อความ ดังสมการ (10) ซึ่งก่อนที่จะคำนวณส่วนนี้ต้องพิจารณาความถี่ของการเกิดร่วมกันระหว่างคำ กับ อารมณ์ของข้อความ เสียก่อนดังตาราง 2.3

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 2.3 ตารางแสดงความถี่ของการเกิดร่วมกันระหว่างคำ กับ อารมณ์ของข้อความ

$n(T_j, E_k)$	ร่าเริง	เสียใจ	โกรธ	กลัว	รังเกียจ	ประหลาดใจ
ไม่	11	185	35	108	71	52
อยาก	28	26	7	9	7	7
น่า	16	52	9	356	16	6
กลัว	0	0	3	441	1	0
เลย	69	53	29	107	27	24
มาก	93	40	12	126	16	10
นี้	27	29	26	68	20	34

$n(T_j, \overline{E_k})$	ร่าเริง	เสียใจ	โกรธ	กลัว	รังเกียจ	ประหลาดใจ
ไม่	1,223	1,049	1,199	1,126	1,163	1,182
อยาก	166	168	187	185	187	187
น่า	569	533	576	229	569	579
กลัว	530	530	527	89	529	530
เลย	575	591	615	537	617	620
มาก	360	413	441	327	437	443
นี้	484	482	485	443	491	477

$n(\overline{T_j}, E_k)$	ร่าเริง	เสียใจ	โกรธ	กลัว	รังเกียจ	ประหลาดใจ
ไม่	825	585	435	1,055	323	473
อยาก	808	744	463	1,154	387	518
น่า	820	718	461	807	378	519
กลัว	836	770	467	722	393	525
เลย	767	717	441	1,056	367	501
มาก	743	730	458	1,037	378	515
นี้	809	741	444	1,095	374	491

$n(\overline{T_j}, \overline{E_k})$	ร่าเริง	เสียใจ	โกรธ	กลัว	รังเกียจ	ประหลาดใจ
ไม่	7,094	7,334	7,484	6,864	7,596	7,446
อยาก	8,151	8,215	8,496	7,805	8,572	8,441
น่า	7,748	7,850	8,107	7,761	8,190	8,049
กลัว	7,787	7,853	8,156	7,901	8,230	8,098
เลย	7,742	7,792	8,068	7,453	8,142	8,008
มาก	7,957	7,970	8,242	7,663	8,322	8,185
นี้	7,833	7,901	8,198	7,547	8,268	8,151

จากตาราง 2.3 เมื่อพิจารณาความถี่ของการเกิดร่วมกันระหว่างของคำ “อยาก” กับอารมณ์ “เสียใจ” จะได้ว่า

$$
\begin{aligned}
n(\text{อยาก}, \text{เสียใจ}) &= 26 \\
n\left(\text{อยาก}, \overline{\text{เสียใจ}}\right) &= 168 \\
n\left(\overline{\text{อยาก}}, \text{เสียใจ}\right) &= 744 \\
n\left(\overline{\text{อยาก}}, \overline{\text{เสียใจ}}\right) &= 8{,}215
\end{aligned}
$$

จากนั้นจะสามารถหาค่า $n(\text{อยาก})$, $n(\overline{\text{อยาก}})$, $n(\text{เสียใจ})$ และ $n\left(\overline{\text{เสียใจ}}\right)$ ได้ดังนี้

$$\begin{aligned}
n(\text{อยาก}) &= n(\text{อยาก},\text{เสียใจ}) + n\left(\text{อยาก},\overline{\text{เสียใจ}}\right) = 26 + 168 = 194 \\
n(\overline{\text{อยาก}}) &= n(\overline{\text{อยาก}},\text{เสียใจ}) + n\left(\overline{\text{อยาก}},\overline{\text{เสียใจ}}\right) = 744 + 8{,}215 = 8{,}959 \\
n(\text{เสียใจ}) &= n(\text{อยาก},\text{เสียใจ}) + n(\overline{\text{อยาก}},\text{เสียใจ}) = 26 + 744 = 770 \\
n\left(\overline{\text{เสียใจ}}\right) &= n\left(\text{อยาก},\overline{\text{เสียใจ}}\right) + n\left(\overline{\text{อยาก}},\overline{\text{เสียใจ}}\right) = 168 + 8{,}215 = 8{,}383
\end{aligned}$$

จำนวนข้อความที่เกิดจากการจับคู่ระหว่างเจ้าของข้อความ กับ ไม่ใช่เจ้าของข้อความ บนข้อมูลชุดฝึกฝน(N) เท่ากับ 9,153 เมื่อแทนค่าในสมการ (10) จะได้ค่าของไคสแควร์ ว่า

$$W_{\chi 2}(\text{อยาก},\text{เสียใจ})$$

$$= \frac{N \times \left[n(\text{อยาก},\text{เสียใจ}) \times n\left(\overline{\text{อยาก}},\overline{\text{เสียใจ}}\right) - n\left(\text{อยาก},\overline{\text{เสียใจ}}\right) \times n(\overline{\text{อยาก}},\text{เสียใจ})\right]^2}{n(\text{อยาก}) \times n(\overline{\text{อยาก}}) \times n(\text{เสียใจ}) \times n\left(\overline{\text{เสียใจ}}\right)}$$

$$= \frac{9{,}153 \times [26 \times 8{,}215 - 168 \times 744]^2}{194 \times 8{,}959 \times 770 \times 8{,}383}$$

$$= 6.404$$

ส่วนที่สาม ค่าของเดลต้า เป็นการคำนวณโดยการวัดคุณลักษณะเฉพาะของคำที่มีต่ออารมณ์ของข้อความ ด้วยการวัดค่าความแตกต่างระหว่าง $n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k})$ กับ $n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k)$ ดังตามสมการที่ 11

$$\delta = \begin{cases} 1, & n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k}) - n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k) > 0 \\ 0, & n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k}) - n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k) \leq 0 \end{cases} \qquad (11)$$

เมื่อพิจารณาคู่ระหว่างคำ “อยาก” กับ อารมณ์ “เสียใจ” จะได้ว่า

$$n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k}) = n(\text{อยาก},\text{เสียใจ}) \times n\left(\overline{\text{อยาก}},\overline{\text{เสียใจ}}\right) = 26 \times 8215 = 213{,}590$$
$$n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k) = n\left(\text{อยาก},\overline{\text{เสียใจ}}\right) \times n(\overline{\text{อยาก}},\text{เสียใจ}) = 168 \times 744 = 124{,}992$$

จะพบว่า $213{,}590 - 124{,}992 > 0$ หรือ $n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k}) - n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k) > 0$

ดังนั้นจะเข้ากรณี $\delta = 1$ นั้นเอง อย่างไรก็ตาม เมื่อพิจารณาคู่ระหว่างคำ “นี้” กับ อารมณ์ “เสียใจ” แทน จะได้ว่า

$n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k}) = n(\text{นี้, เสียใจ}) \times n(\overline{\text{นี้}}, \overline{\text{เสียใจ}}) = 29 \times 7901 = 229{,}129$

$n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k) = n(\text{นี้, เสียใจ}) \times n(\overline{\text{นี้}}, \text{เสียใจ}) = 482 \times 741 = 357{,}162$

จะพบว่า $229{,}129 - 357{,}162 \leq 0$ หรือ $n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k}) - n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k) \leq 0$

ดังนั้นจะเข้ากรณี $\delta = 0$ แทน นั้นคือคำว่า “นี้” ไม่สามารถแยกแยะอารมณ์ “เสียใจ” ได้นั้นเอง

สุดท้าย เมื่อคำนวณทั้งสามส่วนแล้ว นำสามส่วนมาคูณกัน จะได้ค่าน้ำหนักของคำ “อยาก” บนข้อความอารมณ์ “เสียใจ” ดังสมการที่ 12 นั้นคือ

$$W_{tf.\delta-\chi 2}(\text{อยาก, เสียใจ}) = 1 \times \frac{1\,x\,9{,}153 \times [26 \times 8{,}215 - 168 \times 744]^2}{194 \times 8{,}959 \times 770 \times 8{,}383}$$

$$= 6.404$$

2.1.7. การสร้างโมเดลฝึกฝน

การสร้างโมเดลฝึกฝน คือ กระบวนการร้างโมเดลเพื่อให้ระบบได้เรียนรู้การจำแนกอารมณ์ของข้อความ จากข้อมูลชุดฝึกฝน ในงานวิจัยนี้ได้ใช้วิธีการหาฟังก์ชันความหนาแน่นของความน่าจะเป็น สำหรับการกระจายตัวเกาส์เซียนแบบปกติ (Probability Density Function for Normal Gaussian Distribution) [28] ประกอบด้วยพารามิเตอร์ 2 ตัวแปร คือ ค่าเฉลี่ย (Average) กับ ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ของค่าน้ำหนักคำ ดังตามสมการที่ 14

$$P(X = x|E = e) = g(x;\, \mu_e, \sigma_e) = \frac{1}{\sqrt{2\pi}\sigma_e} e^{-\frac{(x-\mu_e)^2}{2{\sigma_e}^2}} \qquad (14)$$

โดยที่ $P(X = x|E = e)$ คือ ความน่าจะเป็นที่ข้อความที่มีคำที่ x บนข้อความอารมณ์ที่ e

$g(x;\, \mu_e, \sigma_e)$ คือ ฟังก์ชันความหนาแน่นของความน่าจะเป็น สำหรับการกระจายตัวเกาส์เซียนแบบปกติ (Probability Density Function for Normal Gaussian Distribution)

μ_e, σ_e คือ ค่าเฉลี่ย กับ ค่าส่วนเบี่ยงเบนมาตรฐาน ของค่าน้ำหนักคำ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.1.8. การจำแนกประเภทของข้อมูล

การจำแนกประเภทของข้อมูล คือ กระบวนการจำแนกหรือทำนายประเภทของข้อความ บนข้อมูลชุดทดสอบ ว่าข้อความดังกล่าวควรอยู่ในอารมณ์ใด โดยพิจารณาแต่ละคำที่อยู่บนข้อความ [10] ในงานวิจัยนี้ได้ใช้วิธีการจำแนกประเภทของข้อมูลด้วยอัลกอริทึมของนาอีฟเบย์ (Naïve Bayes algorithm) [29] ซึ่งประกอบด้วยทฤษฎีเบย์ (Bayes's theorem) [30] กับ แนวคิดว่าคำแต่ละคำมีความเป็นอิสระต่อกัน (Conditionally Independent Assumption) [31]

เริ่มจากการทำนายอารมณ์บนข้อความ จากความน่าจะเป็น $P(E|T)$ ในทางปฏิบัติการหาค่า $P(E|T)$ จะทำได้ยาก และ ใช้ชุดข้อความจำนวนมาก ดังนั้นจึงใช้ทฤษฎีเบย์ (Bayes's theorem) [30] ในการแปลงสมการให้หาค่า $P(T|E)$ แทน ดังตามสมการที่ 15

$$P(E|T) = \frac{P(E) \times P(T|E)}{P(T)} \tag{15}$$

จากเซตของคำ $T = \{t_1, t_2, t_3, \dots, t_M\}$ จะได้ว่า $P(T|E) = P(t_1, t_2, t_3, .., t_M|E)$ และ จากแนวคิดว่าคำแต่ละคำมีความเป็นอิสระต่อกัน (Conditionally Independent Assumption) [31] จึงได้ว่า $P(t_1, t_2, t_3, .., t_M|E) = \prod_{j=1}^{M} P\left(t_j|E\right)$ ดังตามสมการที่ 16

$$P(T|E) = P(t_1, t_2, t_3, .., t_M|E) = \prod_{j=1}^{M} P\left(t_j|E\right) \tag{16}$$

จากสมการที่ 16 แทนลงในสมการ 15 จะได้ตามสมการที่ 17

$$P(E|T) = \frac{P(E) \times \prod_{j=1}^{M} P\left(t_j|E\right)}{P(T)} \tag{17}$$

การทำนายอารมณ์บนข้อความ จะใช้แนวคิด argument maximum (argmax) [29] คือ การหาความน่าจะเป็นของทุกอารมณ์ และ จะเลือกทำนายอารมณ์ที่มีความน่าจะเป็นมากที่สุด ดังตามสมการที่ 18 และ 19

$$e_{predicted} = \underset{e_i \in E}{\operatorname{argmax}}\{P(e_i|T)\} = \underset{e_i \in E}{\operatorname{argmax}}\left\{\frac{P(e_i) \times \prod_{j=1}^{M} P\left(t_j|e_i\right)}{P(T)}\right\} \tag{18}$$

$$e_{NB} = \underset{e_i \in E}{\operatorname{argmax}} \left\{ P(e_i) \prod_{j=1}^{M} P(t_j | e_i) \right\} \quad (19)$$

โดยที่ e_{NB} คือ อารมณ์ที่คำนวณด้วยอัลกรอลิทึมนาอีฟเบย์ [14]

$P(e_i)$ คือ ความน่าจะเป็นที่ข้อความจะเป็นอารมณ์ที่ i

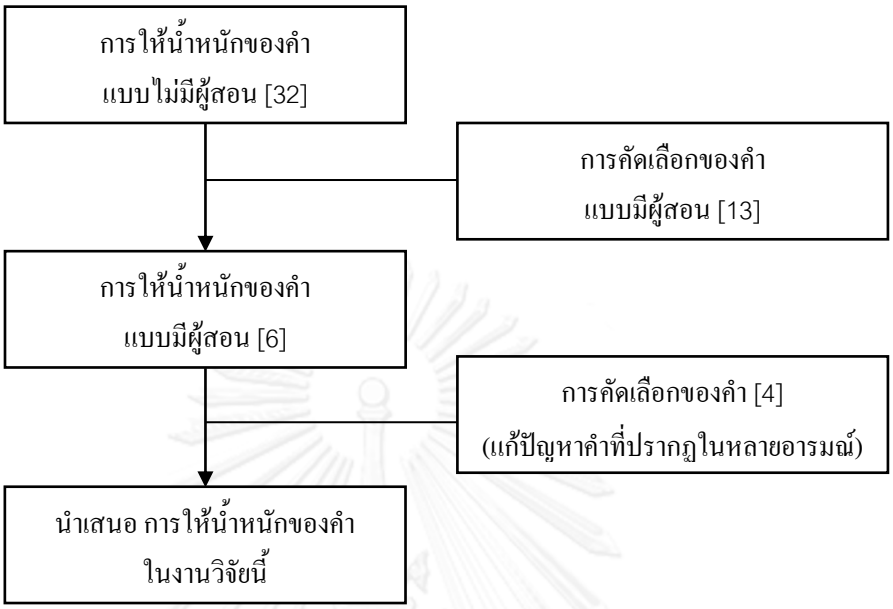
$P(t_j|e_i)$ คือ ความน่าจะเป็นที่ข้อความที่มีคำที่ j บนข้อความอารมณ์ที่ i

M คือ จำนวนรูปแบบของคำ

E คือ เซตของประเภทอารมณ์ 6 อารมณ์ [9]
{ร่าเริง, กลัว, ประหลาดใจ, เสียใจ, รังเกียจ, โกรธ}

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.2. งานวิจัยที่เกี่ยวข้อง



ภาพที่ 2.2 แผนภาพความสัมพันธ์ ของงานวิจัยที่เกี่ยวข้อง

ในงานวิจัยที่เกี่ยวข้องกับการให้น้ำหนักของคำ ในปี ค.ศ. 1988 G. Salton และ C. Buckley [32] ได้ทำการพิจารณาการให้น้ำหนักของคำ ซึ่งใช้บนการจัดเก็บและการค้นคืนสารสนเทศ (Information Retrieval, IR) แบ่งการพิจารณาออกเป็นสามส่วน คือ 1) ความถี่ของคำ (Term Frequency, TF) เป็นค่าที่บ่งบอกความสัมพันธ์ระหว่างคำ กับ ข้อความ โดยผ่านค่าความถี่ของคำที่ปรากฏบนข้อความ 2) อย่างไรก็ตาม การให้น้ำหนักของคำ โดยใช้ความถี่ของคำเพียงอย่างเดียวนั้น ทำให้ค่าน้ำหนักของคำไม่สามารถแยกแยะชุดของข้อความ (Discriminating Term) ที่สัมพันธ์กับชุดข้อความ (Relevant Documents) ออกจากที่ไม่สัมพันธ์กับชุดข้อความ (Irrelevant Documents) ดังนั้นแฟกเตอร์ ส่วนกลับความถี่ของข้อความ (Inverse Document Frequency, IDF) จึงมีบทบาทเพิ่มความสามารถในการแยกแยะชุดของข้อความ ดังนั้นการให้น้ำหนักของคำ จึงประกอบด้วยสองส่วน คือ ความถี่ของคำ กับ ส่วนกลับความถี่ของข้อความ (Term Frequency & Inverse Document Frequency, TF.IDF) 3) ผลกระทบของความยาวข้อความ พิจารณาการปรับค่าน้ำหนักของคำแบบโคซายน์ เพื่อให้แต่ละค่าน้ำหนักของคำอยู่บนบรรทัดฐานเดียวกัน (Cosine Normalization)

ในงานวิจัยที่เกี่ยวข้องกับการคัดเลือกของคำ ในปี ค.ศ. 1997 Y. Yang และ J.O. Pedersen [13] ได้นำเสนอการคัดเลือกของคำ โดยพิจารณาความถี่ในการเกิดร่วมกันระหว่างคำ กับ ข้อความ ได้แก่ Document Frequency (DF), Information Gain (IG), Mutual Information (MI), Chi-Square (CHI) และ Term Strength (TS) ผลจากการทดลองเปรียบเทียบ การคัดเลือกของคำแบบ IG กับ CHI ให้ความแม่นยำมากที่สุด รองลงมา คือ DF, TS และ MI ตามลำดับ จุดประสงค์ของการ

คัดเลือกของคำ คือ การลดขนาดมิติของคำที่มีขนาดใหญ่ (High dimensionality of the Term) [5] โดยการคัดเลือกของคำ ที่สามารถแยกแยะชุดของข้อความ (Discriminating Term) ออกจาก คำที่ไม่สามารถแยกแยะชุดของข้อความได้

ต่อมาในปี ค.ศ. 2003 F. Debole และ F. Sebastiani [6] ได้นำเสนอการให้น้ำหนักของคำโดยการเรียนรู้แบบมีผู้สอน (Supervised Term Weighting) โดยนำแนวคิดการคัดเลือกของคำ มาใช้แทนแฟกเตอร์ ส่วนกลับความถี่ของข้อความ (Inverse Document Frequency, IDF) แล้วนำเสนอการให้น้ำหนักของคำ อันได้แก่ TF.IG, TF.CHI และ TF.GR ผลจากการทดลองเปรียบเทียบโดยวัดค่า macro-averaged F1 ของการให้น้ำหนักของคำ TF.IG, TF.CHI และ TF.GR อยู่ที่ 86%, 79% และ 81% ตามลำดับ บนวิธีการจำแนกข้อความเครื่องจักรเวกเตอร์สนับสนุน (Support Vector Machine, SVM) และในปี ค.ศ. 2004 Z.-H. Deng et al. [32] ได้ทำการเปรียบเทียบการให้น้ำหนักของคำระหว่าง TF.IDF กับ TF.CHI พบว่าการให้น้ำหนักของคำแบบ TF.CHI ให้ประสิทธิภาพที่ดีกว่า TF.IDF โดยค่าของ micro-average F1 ของ TF.CHI อยู่ที่ 87.5%

ในปี ค.ศ. 2008 L. DAI et al. [4] ได้นำเสนอวิธีการคัดเลือกแบบปรับปรุงไคสแควร์ (Modified-Chi Square, δ-$\chi 2$) ผ่านพารามิเตอร์เดลต้า (δ) เป็นวิธีการปรับลดคะแนนของคำที่มีลักษณะเกิดขึ้นในหลากหลายประเภทข้อความ เพื่อให้ได้คำที่เป็นคุณลักษณะเฉพาะที่ชัดเจนขึ้น ได้ผลความแม่นยำอยู่ที่ 86.10% บนวิธีการจำแนกข้อความเครื่องจักรเวกเตอร์สนับสนุน

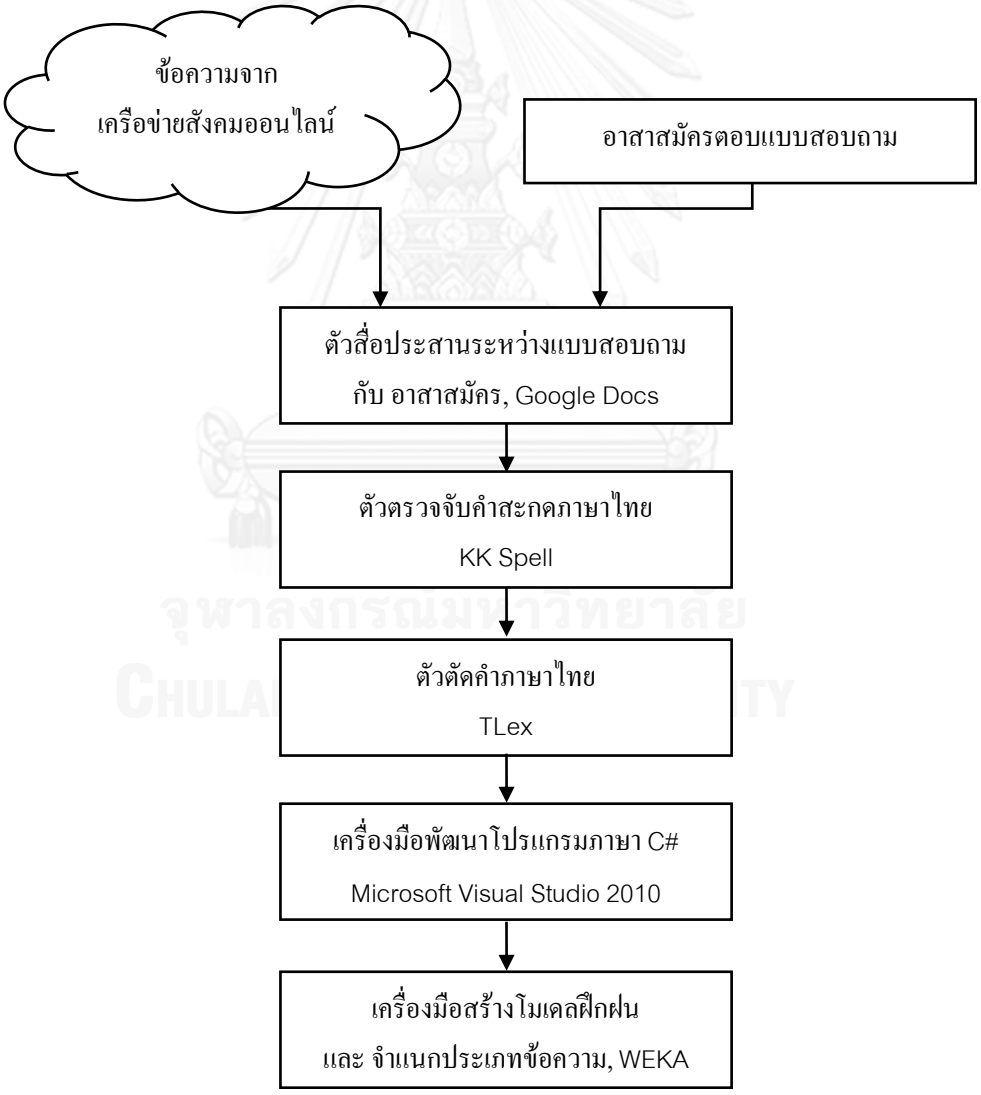
ในส่วนของงานวิจัยนี้ จะมุ่งเน้นนำเสนอการให้น้ำหนักของคำ กับ การคัดเลือกของคำ ในการจำแนกอารมณ์บนข้อความ จากผลการทดลอง ตารางที่ 3.8 พบว่าบางคำมีลักษณะเกิดขึ้นในหลายประเภทอารมณ์ ซึ่งลักษณะคำดังกล่าวมีแนวโน้มจะไม่สามารถแยกแยะชุดของข้อความได้ ดังนั้นเพื่อให้การคัดเลือกของคำ ที่สามารถแยกแยะชุดของข้อความ ออกจาก คำที่ไม่สามารถแยกแยะชุดของข้อความได้ [5] จึงได้นำแนวคิดการปรับลดคะแนนของคำ [4] เพื่อแก้ปัญหาคำที่มีลักษณะเกิดขึ้นในหลากหลายประเภทข้อความ ด้วยการปรับปรุงไคสแควร์ (Modified-Chi Square, δ-$\chi 2$) โดยผ่านพารามิเตอร์เดลต้า (δ) แล้วนำแนวคิดดังกล่าว มาประยุกต์ใช้กับการให้น้ำหนักของคำแบบใหม่ นั้นคือ การให้น้ำหนักของคำแบบความถี่ของคำ กับ เดลต้า-ไคสแควร์ ($tf.\delta$-$\chi 2$)

บทที่ 3
ขั้นตอนการดำเนินงานวิจัย

3. ขั้นตอนการดำเนินงานวิจัย

3.1. เครื่องมือที่ใช้ในการวิจัย

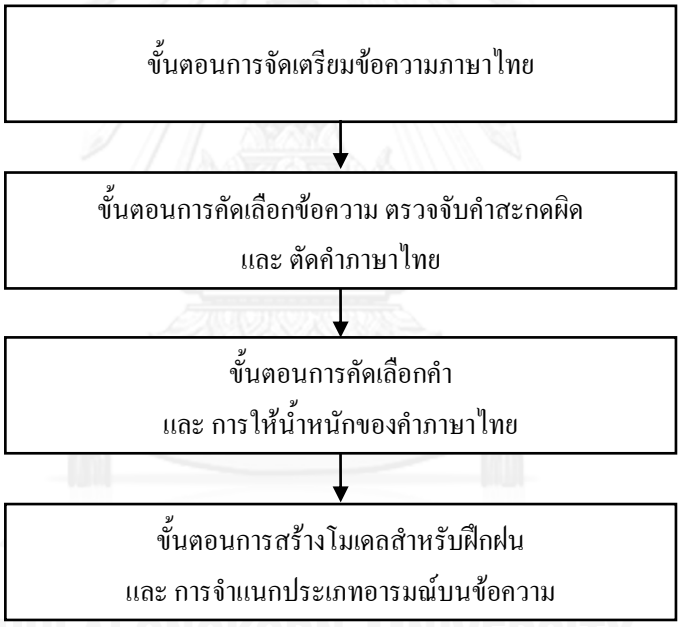
1. ตัวสื่อประสานระหว่างแบบสอบถาม กับ อาสาสมัคร, Google Docs [14]
2. ตัวตรวจจับคำสะกดภาษาไทย, KK Spell [15]
3. ตัวตัดคำภาษาไทย, TLex [12]
4. เครื่องมือพัฒนาโปรแกรมภาษา C#, Microsoft Visual Studio 2010 [16]
5. เครื่องมือสร้างโมเดลฝึกฝน และ จำแนกประเภทข้อความ, WEKA [17]



ภาพที่ 3.1 แผนภาพความสัมพันธ์ของเครื่องมือวิจัย

ในหัวข้อนี้จะพูดถึงเครื่องมือที่ใช้ในงานวิจัยซึ่งการทำงานมีความสัมพันธ์กัน ดังภาพที่ 3.1 ประกอบด้วย 1) ตัวสื่อประสานระหว่างแบบสอบถาม กับ อาสาสมัคร มีหน้าที่สร้างสื่อประสานระหว่างแบบสอบถาม กับ อาสาสมัคร และ ให้อาสาสมัครสามารถตอบแบบสอบถามเพื่อระบุอารมณ์ของข้อความ ข้อความที่รับเข้ามาจะมีทั้งคำที่สะกดถูกต้อง และ คำที่สะกดไม่ถูกต้อง 2) ตัวตรวจจับคำสะกดภาษาไทย มีหน้าที่ตรวจจับคำสะกดภาษาไทยในข้อความนั้น 3) ตัวตัดคำภาษาไทย มีหน้าที่ตัดคำในข้อความภาษาไทย เพื่อวิเคราะห์คุณลักษณะของคำ 4) เครื่องมือพัฒนาโปรแกรมซีชาร์ป มีหน้าที่คำนวณการให้คะแนนคุณลักษณะของคำ เพื่อคัดเลือกคุณลักษณะของคำที่เหมาะสม รวมถึงการให้น้ำหนักของคุณลักษณะของคำในแต่ละข้อความ 5) เครื่องมือสร้างโมเดลฝึกฝน และ จำแนกประเภทข้อความ มีหน้าที่สร้างโมเดลจากชุดข้อมูลฝึกฝน และ การจำแนกประเภทข้อความจากชุดข้อมูลทดสอบ

3.2. ขั้นตอนการดำเนินงานวิจัย



ภาพที่ 3.2 แผนภาพขั้นตอนการดำเนินงานวิจัย

ในหัวข้อนี้จะพูดถึงขั้นตอนการดำเนินงานวิจัย ดังภาพที่ 3.2 ประกอบด้วย 1) ขั้นตอนการจัดเตรียมข้อความภาษาไทย เป็นขั้นตอนจัดเตรียมข้อความจากเครือข่ายสังคมออนไลน์ และ ให้อาสาสมัครระบุอารมณ์ของข้อความ 2) ขั้นตอนการคัดเลือกข้อความ ตรวจจับคำสะกดผิด และ ตัดคำภาษาไทย เป็นขั้นตอนการคัดเลือกชุดข้อความที่เหมาะสมสำหรับฝึกฝนจากการตอบแบบสอบถามของอาสาสมัคร ตรวจจับคำสะกดผิด และ ตัดคำภาษาไทย เพื่อวิเคราะห์คุณลักษณะของคำ 3) ขั้นตอนการคัดเลือกคำ และ การให้น้ำหนักของคำภาษาไทย เป็นขั้นตอนการให้คะแนนคุณลักษณะของคำที่เกิดขึ้นเฉพาะในอารมณ์นั้น แล้วคัดเลือกชุดคุณลักษณะของคำที่เหมาะสม เพื่อ

ทำการให้น้ำหนักของคำ 4) ขั้นตอนการสร้างโมเดลสำหรับฝึกฝน และ การจำแนกประเภทอารมณ์บนข้อความ เป็นขั้นตอนการสร้างโมเดลจากชุดข้อมูลฝึกฝน และ การจำแนกประเภทข้อความจากชุดข้อมูลทดสอบ

3.2.1. ขั้นตอนการจัดเตรียมข้อความภาษาไทย

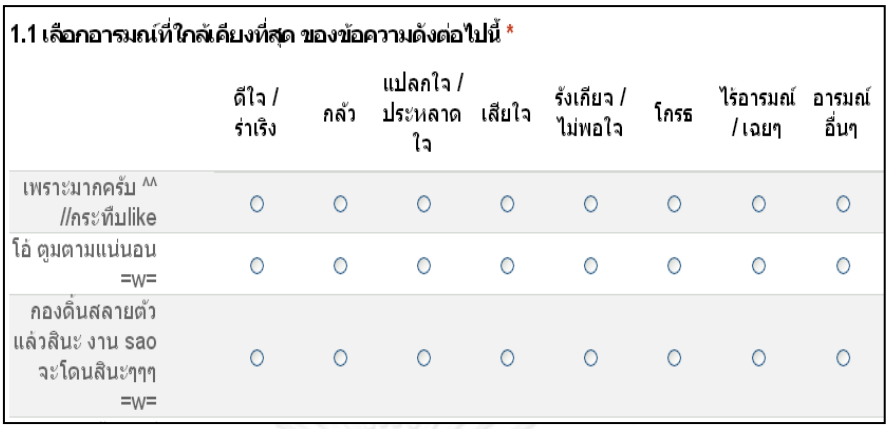
ภาพที่ 3.3 แผนภาพการจัดเตรียมข้อความภาษาไทย

ในหัวข้อนี้จะพูดถึงขั้นตอนการจัดเตรียมข้อความภาษาไทย เริ่มจากข้อความจากเครือข่ายสังคมออนไลน์ ในงานวิจัยนี้ เครือข่ายสังคมออนไลน์ที่สนใจ ได้แก่ เฟสบุ๊ค [7] และ พันทิป [8] ขั้นตอนการดึงข้อความภาษาไทย จะดึงข้อความภาษาไทยที่เจ้าของข้อความเป็นคนโพสในเครือข่ายสังคมออนไลน์ จำนวน 4,609 ข้อความ แบ่งเป็นข้อความจากเฟสบุ๊คจำนวน 1,180 ข้อความ และข้อความจากพันทิปจำนวน 3,429 ข้อความ ขั้นตอนการระบุอารมณ์ของข้อความ จะให้อาสาสมัครทำการระบุอารมณ์ และ คำที่มีลักษณะเด่นเชิงอารมณ์ อารมณ์ในแบบสอบถาม จะประกอบด้วยอารมณ์ร่าเริง, กลัว, ประหลาดใจ, เสียใจ, รังเกียจ, โกรธ, ไม่ปรากฏอารมณ์ และ อารมณ์อื่นๆ อาสาสมัครจะแบ่งออกเป็นสองกลุ่ม คือ 1) กลุ่มเจ้าของข้อความ จะทำการระบุอารมณ์ของข้อความ 2) กลุ่มที่ไม่ใช่เจ้าของข้อความ จะทำการระบุอารมณ์ของข้อความ และ ข้อความที่มีลักษณะเด่นเชิงอารมณ์ ดังภาพที่ 3.4-3.5

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ข้อความที่	แหล่งข้อความ	ตัวอย่างข้อความ
1	พันทิป	แล้วใครจะไปรู้ละครับเมมท่านหูแมวในพันดริ๊บอาจจะชื่อ
2	พันทิป	ใช้ Airnet อยู่ครับ เสถียรดีมากเลยไม่ค่อยเจอปัญหาอะไร
3	พันทิป	นางเอกหายไปทีดิครับคล้ายๆตอนจบของ ef - a tale of melodies
4	พันทิป	เดียวหาอีกไอดีไปเอา ไอดีปัจุบันตัวเลวล 105 คลาส 3 แล้วครับ
5	พันทิป	แต่หลังจาก 5 ทุ่มไปแล้ว วิ่งเต็มๆ เน็ตผม 7 เม๊ก/512 kb
6	พันทิป	คงจะรองตามนี้ก่อน ตอนช่วงที่มันกัดรองเท้าคนนอนหมดแล้วอ่ะครับ
7	พันทิป	ส่วนเมะ BD นั้น มีอะไรดีกว่าที่เคยดูธรรมดาแน่นอนครับ
8	พันทิป	เมนต์ไปว่าทำไมจอ(ที่แอลดู)มันเล็กจัง กร๊ากกกกกกกกกกกก
9	พันทิป	กลัวขึ้นมาหน่อยๆว่าจะโดน emp คอมพังไฟดับน้ำไม่ไหลหรืออย่างไร
10	พันทิป	แล้วทีนี้พอโหลดไปได้ซักเกือบครึ่งทาง มันก็หยุดโหลดอะครับ
11	พันทิป	ทำไมเดาคนตั้งกระทู้และเดาลายฝ่าบ้านได้ถูกวะเนี่ยเรา....
12	พันทิป	ขนาดไม่ได้อวยเมียหลวง ยังรู้สึกอยากตบไอ้แว่นยังไงไม่รู้ =_=
13	พันทิป	ผมก็เป็นครับผม กดโหลดแล้วมันขึ้นว่า "หมดเวลาการเชื่อมต่อ"
14	พันทิป	ไม่ใช่ผมเรียบ หรือหัวมันแหลมๆ ชี้ไปทางโน้นที ทางนั้นที
15	พันทิป	ที่คิริโตะร้องไห้เพราะมันเสียดายฮาเร็มในเกมรึเปล่า
16	เฟสบุ๊ค	ไม่ได้เข้ามานานมากกกก เห็นมี drama อะไรกันใน pantip
17	เฟสบุ๊ค	นิยายนี่มาลงขายในงานหนังสือตั้งแต่วันแรกเลยเปล่าครับ
18	เฟสบุ๊ค	ผมว่า Araragi ทำไปเพราะต้องการช่วยแบบบริสุทธิ์ใจแน่นอน
19	เฟสบุ๊ค	ผมว่าตรงข้ามกันนะ Nise เนี่ยงานภาพดีกว่า Bake เยอะ
20	เฟสบุ๊ค	Krispy Kreme มาไทยแล้วที่พารากอน วันที่ 28 อยากกิน
21	เฟสบุ๊ค	ได้วันหยุดซะที หวัดใหญ่กินซะงั้น อด clear งานที่ค้างเลย
22	เฟสบุ๊ค	ถึงแล้วตั้งแตหกครึ่ง แตเข้าเน็ตไม่ได้เลย เน็ตช้าเกินช้า
23	เฟสบุ๊ค	สงสัยว่าทำไมเค้ากางร่มเร็ว น่าจะรอทำลายสถิติก่อนไหม
24	เฟสบุ๊ค	แต่ตอนนี้ดู Transformer 3 ย้อนหลัง ก็มันส์มากๆเหมือนกัน 5555
25	เฟสบุ๊ค	ฝนตกหนัก ลมกระเชียง เอ๊ย โชกแรง อากาศเย็นน่านอนนนน ^ ^

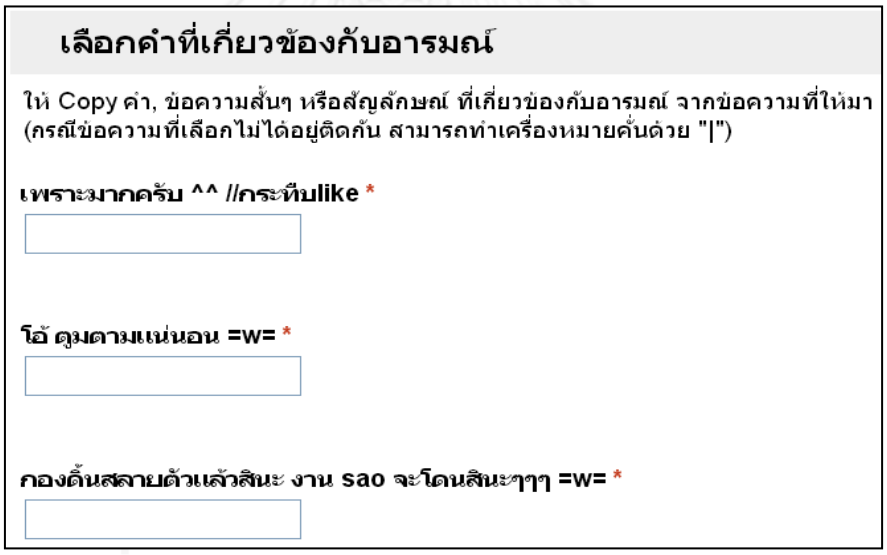
ตารางที่ 3.1 ตารางแสดงตัวอย่างข้อความจากเครือข่ายสังคมออนไลน์



1.1 เลือกอารมณ์ที่ใกล้เคียงที่สุด ของข้อความดังต่อไปนี้ *

	ดีใจ / ร่าเริง	กลัว	แปลกใจ / ประหลาดใจ	เสียใจ	รังเกียจ / ไม่พอใจ	โกรธ	ไร้อารมณ์ / เฉยๆ	อารมณ์อื่นๆ
เพราะมากครับ ^^ //กระทืบlike	○	○	○	○	○	○	○	○
โอ๋ ตูมตามแน่นอน =w=	○	○	○	○	○	○	○	○
กองดิ้นสลายตัวแล้วสินะ งาน sao จะโดนสินะๆๆๆ =w=	○	○	○	○	○	○	○	○

ภาพที่ 3.4 รูปภาพแสดงตัวอย่างแบบสอบถาม ของการระบุอารมณ์ของข้อความ



เลือกคำที่เกี่ยวข้องกับอารมณ์

ให้ Copy คำ, ข้อความสั้นๆ หรือสัญลักษณ์ ที่เกี่ยวข้องกับอารมณ์ จากข้อความที่ให้มา (กรณีข้อความที่เลือกไม่ได้อยู่ติดกัน สามารถทำเครื่องหมายคั่นด้วย "|")

เพราะมากครับ ^^ //กระทืบlike *

โอ๋ ตูมตามแน่นอน =w= *

กองดิ้นสลายตัวแล้วสินะ งาน sao จะโดนสินะๆๆๆ =w= *

ภาพที่ 3.5 รูปภาพแสดงตัวอย่างแบบสอบถาม ของการระบุคำที่มีลักษณะเด่นเชิงอารมณ์

CHULALONGKORN UNIVERSITY

ข้อความ	ตัวอย่างข้อความ	อารมณ์เจ้าของ
1	แล้วใครจะไปรู้ละครับเมมท่านหูแมวในพันดริ๊บอาจจะชื่อ	ประหลาดใจ / แปลกใจ
2	ใช้ Airnet อยู่ครับ เสถียรดีมากเลยไม่ค่อยเจอปัญหาอะไร	ร่าเริง / ดีใจ
3	นางเอกหายไปทีดิครับคล้ายๆตอนจบของ ef - a tale of melodies	ร่าเริง / ดีใจ
4	เดียวหาอีกไอดีไปเอา ไอดีปัจุบันตัวเลวล 105 คลาส 3 แล้วครับ	ร่าเริง / ดีใจ
5	แต่หลังจาก 5 ทุ่มไปแล้ว วิ่งเต็มๆ เน็ตผม 7 เม็ก/512 kb	เสียใจ
6	คงจะรองตามนี้ก่อน ตอนช่วงที่มันกัดรองเท้าคนนอนหมดแล้วอ่ะครับ	เสียใจ
7	ส่วนเมะ BD นั้น มีอะไรดีกว่าที่เคยดูธรรมดาแน่นอนครับ	เฉยๆ / ไร้อารมณ์
8	เมนต์ไปว่าทำไมจอ(ที่แอลดู)มันเล็กจัง กร๊ากกกกกกกกกกก	ประหลาดใจ / แปลกใจ
9	กลัวขึ้นมาหน่อยๆว่าจะโดน emp คอมพังไฟดับน้ำไม่ไหลหรืออย่างไร	กลัว
10	แล้วทีนี้พอโหลดไปได้ซักเกือบครึ่งทาง มันก็หยุดโหลดอะครับ	ประหลาดใจ / แปลกใจ
11	ทำไมเดาคนตั้งกระทู้และเดาลายฝ่าบ้านได้ถูกวะเนี่ยเรา....	ประหลาดใจ / แปลกใจ
12	ขนาดไม่ได้อวยเมียหลวง ยังรู้สึกอยากตบไอ้แว่นยังไงไม่รู้ =_=	ประหลาดใจ / แปลกใจ
13	ผมก็เป็นครับผม กดโหลดแล้วมันขึ้นว่า "หมดเวลาการเชื่อมต่อ"	ประหลาดใจ / แปลกใจ
14	ไม่ใช่ผมเรียบ หรือหัวมันแหลมๆ ชี้ไปทางโน้นที ทางนั้นที	เฉยๆ / ไร้อารมณ์
15	ที่คิริโตะร้องไห้เพราะมันเสียดายฮาเร็มในเกมรึเปล่า	ร่าเริง / ดีใจ
16	ไม่ได้เข้ามานานมากกกก เห็นมี drama อะไรกันใน pantip	เฉยๆ / ไร้อารมณ์
17	นิยายนี่มาลงขายในงานหนังสือตั้งแต่วันแรกเลยเปล่าครับ	เฉยๆ / ไร้อารมณ์
18	ผมว่า Araragi ทำไปเพราะต้องการช่วยแบบบริสุทธิ์ใจแน่นอน	รังเกียจ / ไม่พอใจ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

19	ผมว่าตรงข้ามกันนะ Nise เนี่ยงานภาพดีกว่า Bake เยอะ	รังเกียจ / ไม่พอใจ
20	Krispy Kreme มาไทยแล้วที่พารากอน วันที่ 28 อยากกิน	เฉยๆ / ไร้อารมณ์
21	ได้วันหยุดซะที หวัดใหญ่กินซะงั้น อด clear งานที่ค้างเลย	รังเกียจ / ไม่พอใจ
22	ถึงแล้วตั้งแตหกครึ่ง แตเข้าเน็ตไม่ได้เลย เน็ตช้าเกินช้า	กลัว
23	สงสัยว่าทำไมเค้ากางร่มเร็ว น่าจะรอทำลายสถิติก่อนไหม	ประหลาดใจ / แปลกใจ
24	แต่ตอนนี้ดู Transformer 3 ย้อนหลัง ก็มันส์มากๆเหมือนกัน 5555	ร่าเริง / ดีใจ
25	ฝนตกหนัก ลมกระเชียง เอ๊ย โชกแรง อากาศเย็นน่านอนนนน ^ ^	ร่าเริง / ดีใจ

ตารางที่ 3.2 ตารางแสดงตัวอย่างข้อความ จากการระบุอารมณ์จาก เจ้าของข้อความ

ข้อความที่	**ตัวอย่างข้อความ**	**อารมณ์ที่ไม่ใช่เจ้าของข้อความ**	**ข้อความ ที่มีลักษณะเด่นเชิงอารมณ์**
1	แล้วใครจะไปรู้ละครับเมมท่านหูแมวในพันดริ๊บอาจจะชื่อ	โกรธ	แล้วใครจะไปรู้ละครับ
2	ใช้ Airnet อยู่ครับ เสถียรดีมากเลยไม่ค่อยเจอปัญหาอะไร	ร่าเริง / ดีใจ	เสถียรดีมากเลย
3	นางเอกหายไปทีดิครับคล้ายๆตอนจบของ ef - a tale of melodies	เฉยๆ / ไร้อารมณ์	นางเอกหายไปทีดิครับ
4	เดียวหาอีกไอดีไปเอา ไอดีปัจุบันตัวเลวล 105 คลาส 3 แล้วครับ	เฉยๆ / ไร้อารมณ์	เดียวหาอีกไอดีไปเอา
5	แต่หลังจาก 5 ทุ่มไปแล้ว วิ่งเต็มๆ เน็ตผม 7 เม๊ก/512 kb	ร่าเริง / ดีใจ	5 ทุ่มไปแล้ว วิ่งเต็มๆ เน็ตผม
6	คงจะรองตามนี้ก่อน ตอนช่วงที่มันกัดรองเท้าคนนอนหมดแล้วอ่ะครับ	เฉยๆ / ไร้อารมณ์	คนนอนหมดแล้วอ่ะครับ
7	ส่วนเมะ BD นั้น มีอะไรดีกว่าที่เคยดูธรรมดาแน่นอนครับ	ร่าเริง / ดีใจ	มีอะไรดีกว่าที่เคยดู
8	เมนต์ไปว่าทำไมจอ(ที่แอลดู)มันเล็กจังกร๊ากกกกกกกกกก	ร่าเริง / ดีใจ	กร๊ากกกกกกกกกกกก

9	กลัวขึ้นมาหน่อยๆว่าจะโดน emp คอมพังไฟดับน้ำไม่ไหลหรืออย่างไร	กลัว	กลัวขึ้นมาหน่อยๆว่าจะโดน
10	แล้วทีนี้พอโหลดไปได้ซักเกือบครึ่งทางมันก็หยุดโหลดอะครับ	เสียใจ	มันก็หยุดโหลดอะครับ
11	ทำไมเดาคนตั้งกระทู้และเดาลายฝ่าบ้านได้ถูกวะเนี่ยเรา....	ประหลาดใจ / แปลกใจ	ทำไม
12	ขนาดไม่ได้อวยเมียหลวง ยังรู้สึกอยากตบไอ้แว่นยังไงไม่รู้ =_=	โกรธ	อยากตบไอ้แว่นยังไงไม่รู้
13	ผมก็เป็นครับผม กดโหลดแล้วมันขึ้นว่า "หมดเวลาการเชื่อมต่อ"	เฉยๆ / ไร้อารมณ์	ผมก็เป็นครับผม
14	ไม่ใช่ผมเรียบ หรือหัวมันแหลมๆ ชี้ไปทางโน้นที ทางนั้นที	รังเกียจ / ไม่พอใจ	ไม่ใช่ผมเรียบหรือหัวมันแหลมๆ
15	ที่คิริโตะร้องไห้เพราะมันเสียดายฮาเร็มในเกมรึเปล่า	ประหลาดใจ / แปลกใจ	มันเสียดายฮาเร็มในเกมรึเปล่า
16	ไม่ได้เข้ามานานมากกกก เห็นมี drama อะไรกันใน pantip	ประหลาดใจ / แปลกใจ	เห็นมี drama
17	นิยายนี่มาลงขายในงานหนังสือตั้งแต่วันแรกเลยเปล่าครับ	ประหลาดใจ / แปลกใจ	ตั้งแต่วันแรกเลยเปล่าครับ
18	ผมว่า Araragi ทำไปเพราะต้องการช่วยแบบบริสุทธิ์ใจแน่นอน	เฉยๆ / ไร้อารมณ์	ช่วยแบบบริสุทธิ์ใจ
19	ผมว่าตรงข้ามกันนะ Nise เนี่ยงานภาพดีกว่า Bake เยอะ	ร่าเริง / ดีใจ	ผมว่าตรงข้ามกันนะ
20	Krispy Kreme มาไทยแล้วที่พารากอน วันที่ 28 อยากกิน	ร่าเริง / ดีใจ	อยากกิน
21	ได้วันหยุดซะที หวัดใหญ่กินซะงั้น อด clear งานที่ค้างเลย	เสียใจ	หวัดใหญ่กินซะงั้น
22	ถึงแล้วตั้งแตหกครึ่ง แตเข้าเน็ตไม่ได้เลย เน็ตช้าเกินช้า	เสียใจ	เน็ตช้าเกิน
23	สงสัยว่าทำไมเค้ากางร่มเร็ว น่าจะรอทำลายสถิติก่อนไหม	ประหลาดใจ / แปลกใจ	สงสัยว่าทำไม
24	แต่ตอนนี้ดู Transformer 3 ย้อนหลัง ก็มันส์มากๆเหมือนกัน 5555	ร่าเริง / ดีใจ	ดูครบทุกภาคสินะ
25	ฝนตกหนัก ลมกระเชียง เอ๊ย โชกแรง อากาศเย็นน่านอนนนน ^ ^	ร่าเริง / ดีใจ	อากาศเย็นน่านอนนนน

ตารางที่ 3.3 ตารางแสดงตัวอย่างข้อความ จากการระบุอารมณ์จาก ไม่ใช่เจ้าของข้อความ

3.2.2. ขั้นตอนการคัดเลือกข้อความ ตรวจจับคำสะกดผิด และ ตัดคำภาษาไทย

ภาพที่ 3.6 แผนภาพการคัดเลือกข้อความ ตรวจจับคำสะกดผิด และ ตัดคำภาษาไทย

ในหัวข้อนี้จะพูดถึงขั้นตอนการคัดเลือกข้อความ ตรวจจับคำสะกดผิด และ ตัดคำภาษาไทย เริ่มจากขั้นตอนการคัดเลือกข้อความ จะทำการคัดเลือกข้อความที่มี 6 อารมณ์ [9] ได้แก่ ร่าเริง, กลัว , ประหลาดใจ, เสียใจ, รังเกียจ และ โกรธ จากการโหวตของอาสาสมัครจำนวน 4 คน ในแต่ละข้อความ โดยคัดเลือกข้อความที่มีคะแนนโหวต 3 ใน 4 คนขึ้นไป และ จำนวนอักขระไม่ต่ำกว่า 50 อักขระ โดยจะคัดเลือกข้อความจากการโหวตอารมณ์เหมือนกัน 4 คน และ ข้อความที่มีจำนวนอักขระมากที่สุด เข้ามาก่อน ตามลำดับ ดังแสดงตาราง 3.4.1 และ 3.4.2 ตามลำดับ เพื่อให้ได้ชุดข้อความที่เหมาะสมสำหรับฝึกฝน เพื่อให้ได้ข้อความจาก 6 อารมณ์ที่มีจำนวนเท่ากัน ในงานวิจัยนี้ ได้ทำการคละเรียงอารมณ์บนข้อความจากการโหวต ตัวอย่างเช่น ข้อความที่ 1 ถึง 6 จะเป็นอารมณ์ร่าเริง, กลัว, ประหลาดใจ, เสียใจ, รังเกียจ และ โกรธ ตามลำดับ แล้วข้อความที่ 7 จะทำการเริ่มวนอารมณ์ร่าเริง ขึ้นมาใหม่ วนจนกระทั่งชุดข้อความดังกล่าวไม่ครบทั้ง 6 อารมณ์ จึงทำการหยุดเลือกคัดเลือกข้อความ จำนวนข้อความที่ถูกคัดเลือกจากขั้นตอนนี้ มีจำนวน 3,390 ข้อความ จากทั้งหมด 4,609 ข้อความ ขั้นตอนตรวจจับคำสะกดผิด จะทำการตรวจจับคำสะกดผิดโดยใช้โปรแกรม KK Spell [15] ขั้นตอนการตัดคำภาษาไทย จะทำการแบ่งคำในข้อความภาษาไทย เลือกใช้โปรแกรม TLex [12] โดยใช้อัลกรอลิทึม คอนดิชั่นแนล แรนด้อมฟิลด์ (Conditional Random Fields, CRFs)

ข้อความที่	ตัวอย่างข้อความ	ผลการโหวตอารมณ์
1	แล้วใครจะไปรู้ละครับเมมท่านหูแมวในพันดริ๊บอาจจะชื่อ	รังเกียจ / ไม่พอใจ
2	ใช้ Airnet อยู่ครับ เสถียรดีมากเลยไม่ค่อยเจอปัญหาอะไร	ร่าเริง / ดีใจ
3	นางเอกหายไปทีดิครับคล้ายๆตอนจบของ ef - a tale of melodies	เสียใจ
4	เดียวหาอีกไอดีไปเอา ไอดีปัจุบันตัวเลวล 105 คลาส 3 แล้วครับ	ร่าเริง / ดีใจ
5	แต่หลังจาก 5 ทุ่มไปแล้ว วิ่งเต็มๆ เน็ตผม 7 เม็ก/512 kb	ร่าเริง / ดีใจ
6	คงจะรองตามนี้ก่อน ตอนช่วงที่มันกัดรองเท้าคนนอนหมดแล้วอ่ะครับ	โกรธ
7	ส่วนเมะ BD นั้น มีอะไรดีกว่าที่เคยดูธรรมดาแน่นอนครับ	ร่าเริง / ดีใจ
8	เมนต์ไปว่าทำไมจอ(ที่แอลดู)มันเล็กจัง กร๊ากกกกกกกกก กกก	โกรธ

9	กลัวขึ้นมาหน่อยๆว่าจะโดน emp คอมพังไฟดับน้ำไม่ไหลหรืออย่างไร	กลัว
10	แล้วทีนี้พอโหลดไปได้ซักเกือบครึ่งทาง มันก็หยุดโหลดอะครับ	ประหลาดใจ / แปลกใจ
11	ทำไมเดาคนตั้งกระทู้และเดาลายฝ่าบ้านได้ถูกวะเนี่ยเรา....	ประหลาดใจ / แปลกใจ
12	ขนาดไม่ได้อวยเมียหลวง ยังรู้สึกอยากตบไอ้แว่นยังไงไม่รู้ =_=	โกรธ
13	ผมก็เป็นครับผม กดโหลดแล้วมันขึ้นว่า "หมดเวลาการเชื่อมต่อ"	โกรธ
14	ไม่ใช่ผมเรียบ หรือหัวมันแหลมๆ ชี้ไปทางโน้นที ทางนั้นที	โกรธ
15	ที่คิริโตะร้องไห้เพราะมันเสียดายฮาเร็มในเกมรึเปล่า	เสียใจ
16	ไม่ได้เข้ามานานมากกกก เห็นมี drama อะไรกันในpantip	รังเกียจ / ไม่พอใจ
17	นิยายนี่มาลงขายในงานหนังสือตั้งแต่วันแรกเลยเปล่าครับ	ประหลาดใจ / แปลกใจ
18	ผมว่า Araragi ทำไปเพราะต้องการช่วยแบบบริสุทธิ์ใจแน่นอน	รังเกียจ / ไม่พอใจ
19	ผมว่าตรงข้ามกันนะ Nise เนี่ยงานภาพดีกว่า Bake เยอะ	รังเกียจ / ไม่พอใจ
20	Krispy Kreme มาไทยแล้วที่พารากอน วันที่ 28 อยากกิน	ร่าเริง / ดีใจ
21	ได้วันหยุดซะที หวัดใหญ่กินซะงั้น อด clear งานที่ค้างเลย	ร่าเริง / ดีใจ
22	ถึงแล้วตั้งแตหกครึ่ง แตเข้าเน็ตไม่ได้เลย เน็ตช้าเกินช้า	กลัว
23	สงสัยว่าทำไมเค้ากางร่มเร็ว น่าจะรอทำลายสถิติก่อนไหม	ประหลาดใจ / แปลกใจ
24	แต่ตอนนี้ดู Transformer 3 ย้อนหลัง ก็มันส์มากๆเหมือนกัน 5555	ร่าเริง / ดีใจ
25	ฝนตกหนัก ลมกระเชียง เอ๊ย โชกแรง อากาศเย็นน่านอนนนน ^ ^	ร่าเริง / ดีใจ

ตารางที่ 3.4.1 ตารางแสดงตัวอย่างผลการโหวตอารมณ์ จากอาสาสมัคร

ข้อความที่	ตัวอย่างข้อความ	จำนวนอักขระ
1	แล้วใครจะไปรู้ละครับเมมท่านหูแมวในพันดริ๊บอาจจะชื่อ	51
2	ใช้ Airnet อยู่ครับ เสถียรดีมากเลยไม่ค่อยเจอปัญหาอะไร	53
3	นางเอกหายไปทีดิครับคล้ายๆตอนจบของ ef - a tale of melodies	57
4	เดียวหาอีกไอดีไปเอา ไอดีปัจุบันตัวเลวล 105 คลาส 3 แล้วครับ	59
5	แต่หลังจาก 5 ทุ่มไปแล้ว วิ่งเต็มๆ เน็ตผม 7 เม๊ก/512 kb	54

6	คงจะรองตามนี้ก่อน ตอนช่วงที่มันกัดรองเท้าคนนอนหมดแล้วอ่ะครับ	60
7	ส่วนเมะ BD นั้น มีอะไรดีกว่าที่เคยดูธรรมดาแน่นอนครับ	52
8	เมนต์ไปว่าทำไมจอ(ที่แอลดู)มันเล็กจัง กร๊ากกกกกกกกกกกก	52
9	กลัวขึ้นมาหน่อยๆว่าจะโดน emp คอมพังไฟดับน้ำไม่ไหลหรืออย่างไร	60
10	แล้วทีนี้พอโหลดไปได้ซักเกือบครึ่งทาง มันก็หยุดโหลดอะครับ	57
11	ทำไมเดาคนตั้งกระทู้และเดาลายฝ่าบ้านได้ถูกวะเนี่ยเรา....	54
12	ขนาดไม่ได้อวยเมียหลวง ยังรู้สึกอยากตบไอ้แว่นยังไงไม่รู้ =_=	59
13	ผมก็เป็นครับผม กดโหลดแล้วมันขึ้นว่า "หมดเวลาการเชื่อมต่อ"	57
14	ไม่ใช่ผมเรียบ หรือหัวมันแหลมๆ ชี้ไปทางโน้นที ทางนั้นที	54
15	ที่คิริโตะร้องไห้เพราะมันเสียดายฮาเร็มในเกมรีเปล่า	50
16	ไม่ได้เข้ามานานมากกกก เห็นมี drama อะไรกันใน pantip	51
17	นิยายนี่มาลงขายในงานหนังสือตั้งแต่วันแรกเลยเปล่าครับ	52
18	ผมว่า Araragi ทำไปเพราะต้องการช่วยแบบบริสุทธิ์ใจแน่นอน	54
19	ผมว่าตรงข้ามกันนะ Nise เนี่ยงานภาพดีกว่า Bake เยอะ	50
20	Krispy Kreme มาไทยแล้วที่พารากอน วันที่ 28 อยากกิน	50
21	ได้วันหยุดซะที หวัดใหญ่กินซะงั้น อด clear งานที่ค้างเลย	55
22	ถึงแล้วตั้งแตหกครึ่ง แตเข้าเน็ตไม่ได้เลย เน็ตช้าเกินช้า	55
23	สงสัยว่าทำไมเค้ากางร่มเร็ว น่าจะรอทำลายสถิติก่อนไหม	51
24	แต่ตอนนี้ดู Transformer 3 ย้อนหลัง ก็มันส์มากๆเหมือนกัน 5555	60
25	ฝนตกหนัก ลมกระเชียง เอ๊ย โชกแรง อากาศเย็นน่านอนนนน ^ ^	54

ตารางที่ 3.4.2 ตารางแสดงตัวอย่าง และ จำนวนอักขระ

ข้อความที่	**ตัวอย่างข้อความที่ตรวจจับ และ แก้ไขคำสะกดผิด**
1	แล้วใครจะไปรู้ละครับเมท่านหูแมวในพันทิปอาจจะชื่อ
2	ใช้ Airnet อยู่ครับ เสถียรดีมากเลยไม่ค่อยเจอปัญหาอะไร
3	นางเอกหายไปทีดิครับคล้าย ตอนจบของ ef a tale of melodies
4	เดียวหาอีกไอดีไปเอา ไอดีปัจจุบันตัวเลวล 105 คลาส 3 แล้วครับ
5	แต่หลังจาก 5 ทุ่มไปแล้ว วิ่งเต็ม เน็ตผม 7เมก512kb
6	คงจะรองตามนี้ก่อน ตอนช่วงที่มันกัดรองเท้าคนนอนหมดแล้วอ่ะครับ
7	ส่วน เมะBD นั้น มีอะไรดีกว่าที่เคยดูธรรมดาแน่นอนครับ
8	เมนต์ไปว่าทำไมจอ ที่แอลดู มันเล็กจัง กร๊าก
9	กลัวขึ้นมาหน่อย ว่าจะโดน emp คอมพังไฟดับน้ำไม่ไหลหรืออย่างไร
10	แล้วทีนี้พอโหลดไปได้ซักเกือบครึ่งทาง มันก็หยุดโหลดอะครับ

11	ทำไมเดาคนตั้งกระทู้และเดาลายฝ่าบ้านได้ถูกวะเนี่ยเรา
12	ขนาดไม่ได้อวยเมียหลวง ยังรู้สึกอยากตบไอ้แว่นยังไงไม่รู้=_=
13	ผมก็เป็นครับผม กดโหลดแล้วมันขึ้นว่า หมดเวลาการเชื่อมต่อ
14	ไม่ใช่ผมเรียบ หรือหัวมันแหลม ชี้ไปทางโน้นที ทางนั้นที
15	ที่คิริโตะร้องไห้เพราะมันเสียดายฮาเร็มในเกมรึเปล่า
16	ไม่ได้เข้ามานานมาก เห็นมี drama อะไรกันใน pantip
17	นิยายนี่มาลงขายในงานหนังสือตั้งแต่วันแรกเลยเปล่าครับ
18	ผมว่า Araragi ทำไปเพราะต้องการช่วยแบบบริสุทธิ์ใจแน่นอน
19	ผมว่าตรงข้ามกันนะ Nise เนี่ยงานภาพดีกว่า Bake เยอะ
20	Krispy Kreme มาไทยแล้วที่พารากอน วันที่ 28 อยากกิน
21	ได้วันหยุดซะที หวัดใหญ่กินซะงั้น อด clear งานที่ค้างเลย
22	ถึงแล้วตั้งแต่หกครึ่ง แต่เข้าเน็ตไม่ได้เลย เน็ตช้าเกินช้า
23	สงสัยว่าทำไมเค้ากางร่มเร็ว น่าจะรอทำลายสถิติก่อนไหม
24	แต่ตอนนี้ดู Transformer 3 ย้อนหลัง ก็มันส์มาก เหมือนกัน ฮ่า
25	ฝนตกหนัก ลมกระเชียง เอ๊ย โชกแรง อากาศเย็นน่านอน ^_^
26	เทศกาลกินเจ เมล่อนที่รักของหนูต้องราคาขึ้นแน่
27	โอ๊ยเครียด รูมเมทซื้อทุเรียนมากิน ตรูหนีไปไหนไม่ได้
28	ไม่มีคำว่าช้าไปที่จะรู้จัก ไม่มีคำว่าเร็วไปที่จะรักเธอ
29	สั่งตัวเองว่าต้องเลิกฟุ้งซ่านแล้วทำ thesis ต่อให้เสร็จ
30	เผื่อไปอะไรกะใครได้ยื่นบัตรก่อนพี่ดรีมเหอ เดะถูกฉัด

ตารางที่ 3.5 ตารางแสดงตัวอย่างการตรวจจับ และ แก้ไขคำสะกดผิด

ข้อความที่	**ตัวอย่างการตัดคำบนข้อความภาษาไทย**
1	แล้ว|ใคร|จะ|ไป|รู้|ละ|ครับ|เมม|ท่าน|หู|แมว|ใน|พันทิป|อาจ|จะ|ชื่อ|
2	ใช้| |Airnet| |อยู่|ครับ| |เสถียร|ดี|มาก|เลย|ไม่|ค่อย|เจอ|ปัญหา|อะไร|
3	นางเอก|หาย|ไป|ที|ดิ|ครับคล้าย| || |ตอน|จบ|ของ| |ef| || |a| |tale| |of| |melodies|
4	เดียวหา|อีก|ไอดี|ไป|เอา| |ไอดี|ปัจจุบัน|ตัว|เลวล| |105| |คลาส| |3| |แล้ว|ครับ|
5	แต่|หลัง|จาก| |5| |ทุ่ม|ไป|แล้ว| |วิ่ง|เต็ม| || | |เน็ต|ผม| |7|เมก||512|kb|
6	คง|จะ|รอง|ตาม|นี้|ก่อน| |ตอน|ช่วง|ที่|มัน|กัด|รองเท้า|คน|นอน|หมด|แล้ว|อ่ะ|ครับ|
7	ส่วน| |เมะ|BD| |นั้น| |มี|อะไร|ดี|กว่า|ที่|เคย|ดู|ธรรมดา|แน่นอน|ครับ|
8	เมนต์|ไป|ว่า|ทำไมจอ| |ที่|แอลดู| |มัน|เล็ก|จัง| |กร๊าก|||
9	กลัว|ขึ้น|มา|หน่อย| || |ว่า|จะ|โดน| |emp| |คอมพัง|ไฟ|ดับ|น้ำ|ไม่|ไหล|หรือ|อย่าง|ไร|
10	แล้ว|ทีนี้|พอ|โหลด|ไป|ได้|ซัก|เกือบ|ครึ่ง|ทาง| |มัน|ก็|หยุด|โหลด|อะ|ครับ|
11	ทำไม|เดา|คน|ตั้ง|กระทู้|และ|เดา|ลาย|ฝา|บ้าน|ได้|ถูก|วะ|เนี่ย|เรา||

12	ขนาด|ไม่|ได้|อวย|เมีย|หลวง| |ยัง|รู้สึก|อยาก|ตบ|ไอ้|แว่น|ยัง|ไง|ไม่|รู้|=_=|
13	ผม|ก็|เป็น|ครับ|ผม| |กด|โหลด|แล้ว|มัน|ขึ้น|ว่า| ||หมด|เวลาการ|เชื่อม|ต่อ||
14	ไม่|ใช่|ผม|เรียบ| |หรือ|หัว|มัน|แหลม| || | |ชี้|ไป|ทาง|โน้น|ที| |ทาง|นั้น|ที|
15	ที่|คิริโตะ|ร้องไห้|เพราะ|มัน|เสียดายฮาเร็ม|ใน|เกม|รึ|เปล่า|
16	ไม่|ได้|เข้า|มา|นาน|มาก| |เห็น|มี| |drama| |อะไร|กัน|ใน| |pantip|
17	นิยาย|นี่|มา|ลง|ขาย|ใน|งาน|หนังสือ|ตั้งแต่|วัน|แรก|เลย|เปล่า|ครับ|
18	ผม|ว่า| |Araragi| |ทำไป|เพราะ|ต้องการ|ช่วย|แบบ|บริสุทธิ์ใจ|แน่นอน|
19	ผม|ว่า|ตรงข้าม|กัน|นะ| |Nise| |เนี่ย|งานภาพ|ดี|กว่า| |Bake| |เยอะ|
20	Krispy| |Kreme| |มา|ไทย|แล้ว|ที่|พารากอน| |วัน|ที่| |28| |อยาก|กิน|
21	ได้|วัน|หยุด|ซะ|ที| |หวัด|ใหญ่|กิน|ซะ|งั้น| |อด| |clear| |งาน|ที่|ค้าง|เลย|
22	ถึง|แล้ว|ตั้งแต่|หก|ครึ่ง| |แต่|เข้าเน็ต|ไม่ได้|เลย| |เน็ต|ช้า|เกิน|ช้า|
23	สงสัย|ว่า|ทำไม|เค้า|กาง|ร่ม|เร็ว| |น่า|จะ|รอ|ทำลาย|สถิติ|ก่อน|ไหม|
24	แต่|ตอน|นี้|ดู| |Transformer| |3| |ย้อน|หลัง| |ก็|มันส์|มาก| || |เหมือน|กัน| |ฮ่า|
25	ฝน|ตก|หนัก| |ลม|กระเชียง| |เอ๊ย| |โชก|แรง| |อากาศ|เย็น|น่า|นอน| |^_^|
26	เทศกาล|กินเจ| || |เมล่อน|ที่|รัก|ของ|หนู|ต้อง|ราคา|ขึ้น|แน่| ||
27	โอ๊ย|เครียด| |รูมเมท|ซื้อ|ทุเรียน|มา|กิน| |ตรู|หนี|ไป|ไหน|ไม่|ได้|
28	ไม่|มี|คำ|ว่า|ช้า|ไป|ที่|จะ|รู้จัก| |ไม่|มี|คำ|ว่า|เร็ว|ไป|ที่|จะ|รัก|เธอ|
29	สั่ง|ตัว|เอง|ว่า|ต้อง|เลิก|ฟุ้งซ่าน|แล้ว|ทำ| |thesis| |ต่อให้|เสร็จ|
30	เผื่อ|ไป|อะไร|กะ|ใคร|ได้|ยื่น|บัตร|ก่อน|พี่|ดรีม|เหอ| |เดะ|ถูก|ฉัด|

ตารางที่ 3.6 ตารางแสดงตัวอย่างการตัดคำบนข้อความภาษาไทย

3.2.3. ขั้นตอนการคัดเลือก และ การให้น้ำหนักของคำภาษาไทย

ภาพที่ 3.7 แผนภาพการการคัดเลือก และ การให้น้ำหนักของคำภาษาไทย

ในหัวข้อนี้จะพูดถึงขั้นตอนคัดเลือก และ การให้น้ำหนักของคำภาษาไทย เริ่มจากการสร้างชุดของคำที่สะกดไม่ซ้ำกันในชุดข้อมูลฝึก ได้จำนวนทั้งหมด 4,260 คำ ขั้นตอนการนับความถี่การเกิดร่วมกันระหว่างคำ กับ อารมณ์ [13] จะทำการนับจำนวนที่เกิดร่วมกันระหว่างคำ กับ อารมณ์ของข้อความ โดยพิจารณาคำ และ อารมณ์ของข้อความ ที่เข้าคู่กันระหว่างอาสาสมัครที่เป็นเจ้าของข้อความ และ ไม่ใช่เจ้าของข้อความ ขั้นตอนการให้คะแนนของคำ จะทำการให้คะแนนของคำในชุด

ข้อมูลฝึก โดยใช้การคำนวณด้วยวิธีเดลต้าไคสแควร์ (Delta-Chi Square, δ-$\chi 2$) [4] ซึ่งค่าของคะแนนจะขึ้นอยู่กับคำที่เกิดขึ้นเฉพาะในอารมณ์นั้น ขั้นตอนการเรียงคะแนน และ การคัดเลือกของคำ จะทำการเรียงคะแนนจากมากไปน้อย และ คะแนนที่มากกว่าจะถูกคัดเลือกคำออกมาก่อน ในงานวิจัยนี้ จะคัดเลือกคำจำนวน 10, 40, 160, 640, 2,560 และ 4,260 คำ ขั้นตอนการสร้างแบบจำลองปริภูมิเวกเตอร์ [1] จะทำการแปลงจากข้อความเป็นแบบจำลองปริภูมิเวกเตอร์ เพื่อให้ระบบวิเคราะห์คำในข้อความนั้นได้ ขั้นตอนการคำนวณค่าน้ำหนักของคำ จะทำการคำนวณค่าน้ำหนักของคำโดยใช้การคำนวณด้วยวิธีนับความถี่ของคำ กับ เดลต้า-ไคสแควร์ (Term Frequency Delta-Chi Square, $tf.\delta$-$\chi 2$) [4]

คำที่	ตัวอย่างคำ
1	ไม่
2	น่า
3	กลัว
4	เลย
5	มาก
6	ไป
7	ได้
8	แล้ว
9	นี้
10	มัน

คำที่	ตัวอย่างคำ
11	จะ
12	ดู
13	ก็
14	มา
15	ครับ
16	คน
17	มี
18	นะ
19	หลอน
20	ที่

คำที่	ตัวอย่างคำ
21	กัน
22	แต่
23	อะไร
24	ดี
25	จริง
26	ผม
27	ทำ
28	เป็น
29	อ่ะ
30	จัง

ตารางที่ 3.7 ตารางแสดงตัวอย่างการสร้างชุดของคำ บนชุดข้อมูลฝึกฝน

ตัวอย่างคำ / อารมณ์	ร่าเริง / ดีใจ	เสียใจ	โกรธ	กลัว	รังเกียจ / ไม่พอใจ	ประหลาดใจ / แปลกใจ
ไม่	11	185	35	108	71	52
น่า	16	52	9	356	16	6
กลัว	0	0	3	441	1	0
เลย	69	53	29	107	27	24
มาก	93	40	12	126	16	10
ไป	30	57	38	32	44	30
ได้	42	77	22	14	27	42
แล้ว	42	47	19	70	19	24
นี้	27	29	26	68	20	34
มัน	16	13	40	28	33	62

ศูนย์วิทยทรัพยากร
Chulalongkorn University

จะ	24	21	33	45	35	26
ดู	16	33	6	93	19	12
ก็	32	40	14	48	23	11
มา	56	28	19	29	12	22
ครับ	55	21	14	18	13	44
คน	38	38	25	27	19	14
มี	28	30	5	26	26	29
นะ	36	20	9	33	11	18
หลอน	0	0	0	125	0	0
ที่	27	29	12	22	24	6
กัน	20	18	28	16	12	24
แต่	14	39	9	18	16	8
อะไร	1	0	18	6	20	54
ดี	58	3	9	10	15	3
จริง	29	5	16	26	1	21
ผม	29	17	8	17	17	9
ทำ	12	16	23	5	16	23
เป็น	18	25	6	10	12	23
อ่ะ	10	14	8	43	5	12
จัง	5	36	4	45	0	0

ตารางที่ 3.8 ตารางแสดงตัวอย่างการนับความถี่การเกิดร่วมกันระหว่างคำ กับ อารมณ์
โดยการเรียนรู้จากการจับคู่ระหว่างคนที่เป็นเจ้าของข้อความ และ คนที่ไม่เป็นเจ้าของข้อความ

ตัวอย่างคำ	คะแนน
ไม่	7.05
น่า	165.98
กลัว	320.32
เลย	1.4
มาก	19.28
ไป	0.8
ได้	1.8
แล้ว	0.06
นี้	0.07
มัน	2.66

ตัวอย่างคำ	คะแนน
จะ	0.28
ดู	7.97
ก็	0.27
มา	0.55
ครับ	1.63
คน	0
มี	0.36
นะ	0.19
หลอน	99.21
ที่	0.33

ตัวอย่างคำ	คะแนน
กัน	0.83
แต่	3.11
อะไร	9.07
ดี	7.73
จริง	1.11
ผม	0.45
ทำ	0.73
เป็น	0.09
อ่ะ	5.15
จัง	10.31

ตารางที่ 3.9 ตารางแสดงตัวอย่างการให้คะแนนของคำ แบบเดลต้า-ไคสแควร์

ลำดับที่	ตัวอย่างคำ	คะแนน	ลำดับที่	ตัวอย่างคำ	คะแนน	ลำดับที่	ตัวอย่างคำ	คะแนน
1	กลัว	320.32	11	ผี	26.57	21	มาก	19.28
2	น่า	165.98	12	เจ็บ	26.06	22	ขอบคุณ	19.25
3	หลอน	99.21	13	เศร้า	24.14	23	คืน	16.44
4	ฮ่า	39.05	14	สยอง	23.20	24	ไอ้	15.96
5	ขนลุก	34.38	15	เสียใจ	22.83	25	สนุก	14.12
6	T_T	32.68	16	ตกใจ	22.23	26	หรอ	11.79
7	สงสาร	32.47	17	ฮา	22.13	27	หว่า	11.41
8	ขน	30.97	18	ร้องไห้	20.95	28	ล่า	10.88
9	กล้า	30.37	19	เหรอ	20.43	29	จัง	10.31
10	ลุก	30.10	20	เสียว	19.32	30	คร้าบ	10.04

ตารางที่ 3.10 ตารางแสดงตัวอย่างการเรียงคะแนนของคำ แบบเดลต้า-ไคสแควร์

ตัวอย่างข้อความ	ค่าน้ำหนักของคำ						
	ไม่	ไป	มัน	ได้	จะ	กลัว	...
อยาก|ดู|แต่|ดู|ไม่|ได้|เพราะ|พรุ่ง|นี้|ต้อง|ไป|โรง|เรียน|	W_{11}	W_{12}	W_{13}	W_{14}	W_{15}	W_{16}	...
กู|จะ|ตาม|ดู|ไป|จนกว่า|มัน|จะ|เลิก|สร้าง|ภาค|ต่อ|เลย|	W_{21}	W_{22}	W_{23}	W_{24}	W_{25}	W_{26}	...
ดู|ไป|ดู|มา|เริ่ม|ไม่|ฮา|ละ| |น่า|สงสาร|จิง| |=w=|	W_{31}	W_{32}	W_{33}	W_{34}	W_{35}	W_{36}	...
ไม่|น่า|พลาด|ไป|ดู|มัน|ทั้ง|วัน|เลย| |แค่|นั้น|แหละ|	W_{41}	W_{42}	W_{43}	W_{44}	W_{45}	W_{46}	...
มัน|ซัก|จะ|ไม่|ใช่|โรง|เรียน|เข้า|ไป|ทุก|ที|แล้ว| |-_-||	W_{51}	W_{52}	W_{53}	W_{54}	W_{55}	W_{56}	...
อยาก|ได้| |อยาก|ไป|ดู| |แต่|อยู่| |ตจว| |ค่ะ| |T_T| |ฮือ|	W_{61}	W_{62}	W_{63}	W_{64}	W_{65}	W_{66}	...
อยาก|ได้|มา|อวย|บ้าง| |แต่|หา|ซื้อ|ไม่|ค่อย|ได้|แล้ว|	W_{71}	W_{72}	W_{73}	W_{74}	W_{75}	W_{76}	...
อยาก|ไป|อีก|แต่|ไม่|ค่อย|ว่าง|แล้ว| |เสียดาย|เหมือน|กัน|	W_{81}	W_{82}	W_{83}	W_{84}	W_{85}	W_{86}	...

ไม่|น่า|กลัว|เท่า|ไหร่| |แต่|ก็|เสียว|สยอง|ลึก||เหมือน|กัน|	W_{91}	W_{92}	W_{93}	W_{94}	W_{95}	W_{96}	...
Elfen| |lied| |ครับ| |น่า|กลัว|จริง|| |แต่|ได้|อารมณ์|มาก||	W_{101}	W_{102}	W_{103}	W_{104}	W_{105}	W_{106}	...

ตารางที่ 3.11 ตารางแสดงตัวอย่างการสร้างแบบจำลองปริภูมิเวกเตอร์

ตัวอย่างข้อความ	**ค่าน้ำหนักของคำ**						
	ไม่	ไป	มัน	ได้	จะ	กลัว	...
อยาก|ดู|แต่|ดู|ไม่|ได้|เพราะ|พรุ่ง|นี้|ต้อง|ไป|โรง|เรียน|	0.827	0.010	0	0.201	0	0	...
กู|จะ|ตาม|ดู|ไป|จนกว่า|มัน|จะ|เลิก|สร้าง|ภาค|ต่อ|เลย|	0	0.095	0.327	0	0.077	0	...
ดู|ไป|ดู|มา|เริ่ม|ไม่|ฮา|ละ| |น่า|สงสาร|จิง| |=w=|	0.203	0.003	0	0	0	0	...
ไม่|น่า|พลาด|ไป|ดู|มัน|ทั้ง|วัน|เลย| |แค่|นั้น|แหละ|	0.768	0.010	0	0	0	0	...
มัน|ซัก|จะ|ไม่|ใช่|โรง|เรียน|เข้า|ไป|ทุก|ที|แล้ว| |-_-||	0	0.275	0.951	0	0.111	0	...
อยาก|ได้| |อยาก|ไป|ดู| |แต่|อยู่| |ตจว| |ค่ะ| |T_T| |ฮือ|	0	0.003	0	0.049	0	0	...
อยาก|ได้|มา|อวย|บ้าง| |แต่|หา|ซื้อ|ไม่|ค่อย|ได้|แล้ว|	0.200	0	0	0.097	0	0	...
อยาก|ไป|อีก|แต่|ไม่|ค่อย|ว่าง|แล้ว| |เสียดาย|เหมือน|กัน|	0.664	0.008	0	0	0	0	...
ไม่|น่า|กลัว|เท่า|ไหร่| |แต่|ก็|เสียว|สยอง|ลึก||เหมือน|กัน|	0	0	0	0	0	0.885	...
Elfen| |lied| |ครับ| |น่า|กลัว|จริง|| |แต่|ได้|อารมณ์|มาก||	0	0	0	0	0	0.887	...

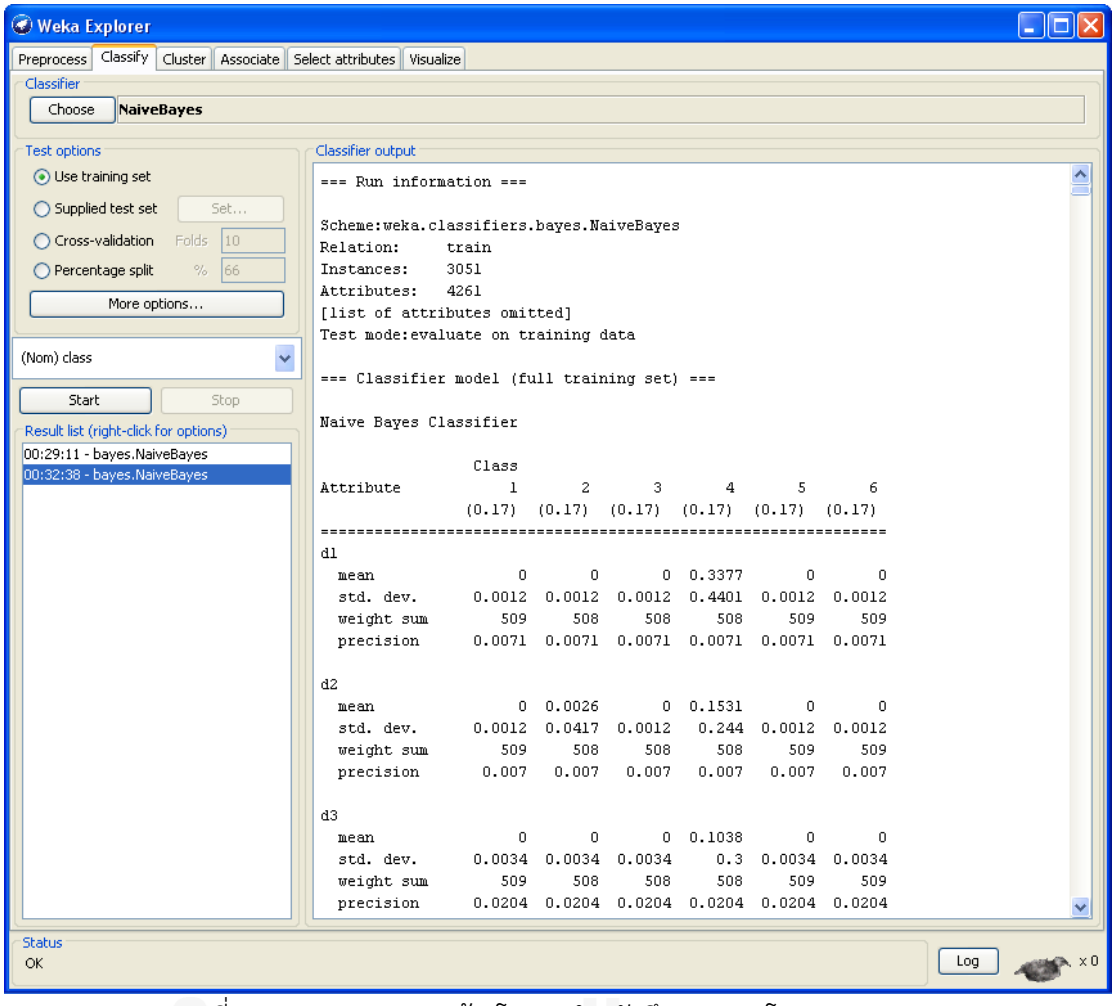
ตารางที่ 3.12 ตารางแสดงตัวอย่างการคำนวณค่าน้ำหนักของคำ

3.2.4. ขั้นตอนการสร้างโมเดลฝึกฝน และ การจำแนกประเภทอารมณ์บนข้อความ

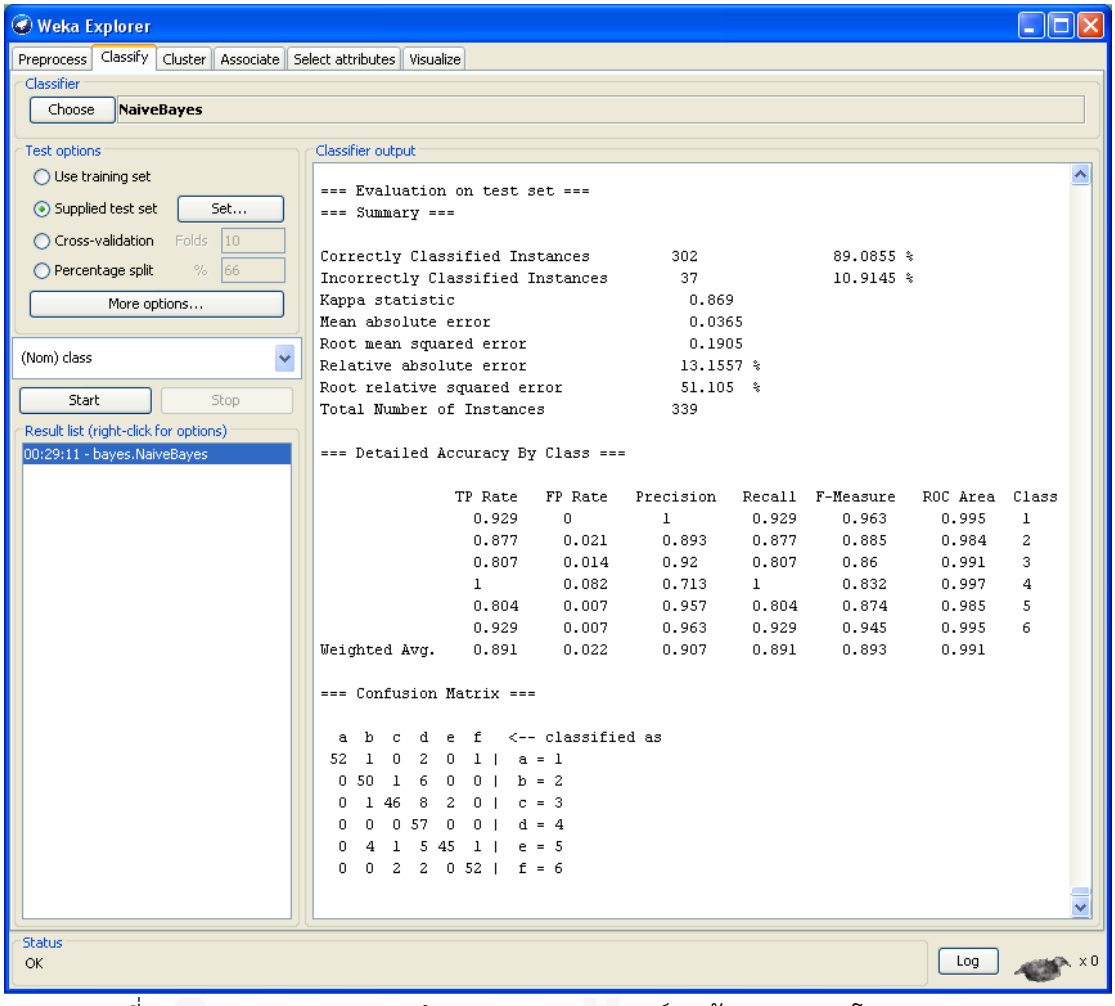
ภาพที่ 3.8 แผนภาพการสร้างโมเดลสำหรับฝึกฝน และ การจำแนกประเภทอารมณ์บนข้อความ

ในหัวข้อนี้จะพูดถึงขั้นตอนการสร้างโมเดลสำหรับฝึกฝน และ การจำแนกประเภทอารมณ์บนข้อความ เริ่มจากการสร้างโมเดลสำหรับฝึกฝน จะทำการหาค่าเฉลี่ย (Mean) กับ ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ของค่าน้ำหนักของคำ กับ อารมณ์ เพื่อนำมาสร้างโมเดลการกระจายเกาส์เซียนแบบปกติ (Normal Gaussian Distribution) [10] ขั้นตอนการจำแนกประเภทอารมณ์บนข้อความ จะจำแนกประเภทข้อความโดยใช้อัลกรอลิทึมการจำแนกแบบนาอีฟเบย์ (Naïve Bayes Classification) [10] เป็นอัลกรอลิทึมการจำแนกที่ประยุกต์จากทฤษฎีเบย์ (Bayes's Theorem) โดยใช้สมมุติฐานของคำแต่ละคำมีอิสระต่อกัน (Independence Term Assumption

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



ภาพที่ 3.9 ภาพแสดงการสร้างโมเดลสำหรับฝึกฝน ของโปรแกรม WEKA



ภาพที่ 3.10 ภาพแสดงการการจำแนกประเภทอารมณ์บนข้อความ ของโปรแกรม WEKA

บทที่ 4
การทดลอง และอภิปรายผล

4. การทดลอง และอภิปรายผล

ในงานวิจัยนี้ ข้อความที่นำเข้ามาจากเครือข่ายสังคมออนไลน์ คือ เฟสบุ๊ค กับ พันทิป เป็นจำนวน 4,609 ข้อความ คัดเลือกข้อความจากการโหวตอารมณ์ของอาสาสมัคร เหลือจำนวน 3,390 ข้อความ ประกอบด้วย ข้อความมาจากพันทิป จำนวน 2,502 ข้อความ และ ข้อความมาจากเฟสบุ๊ค จำนวน 888 ข้อความ ในจำนวน 3,390 ข้อความ แบ่งเป็นอารมณ์ของข้อความ อารมณ์ละ 565 ข้อความ แล้วทำการแบ่งเป็นชุดข้อมูลสำหรับฝึกฝน จำนวน 3,051 ข้อความ และ แบ่งเป็นชุดข้อมูลสำหรับทดสอบ จำนวน 339 ข้อความ

ในการทดลองระบบการจำแนกอารมณ์บนข้อความ จากเครือข่ายสังคมออนไลน์ ได้แบ่งการทดลองออกเป็น 4 การทดลองด้วยกัน ได้แก่ 1) การทดสอบแบบไขว้ข้าม (K-Cross Validation) 2) การทดสอบความแม่นยำในการจำแนกอารมณ์บนข้อความ (Accuracy) 3) การทดลองความแม่นยำของวิธีการคัดเลือกของคำ เปรียบเทียบระหว่างวิธีการคัดเลือกตามหน้าที่คำ (Part of Speech, POS) กับ คัดเลือกโดยให้คะแนนแล้วเรียงคะแนน 4) การทดลองเปรียบเทียบความแม่นยำบนชุดข้อความระหว่างเฟสบุ๊ค กับ พันทิป

4.1. การทดสอบแบบไขว้ข้าม (Cross Validation)

การทดสอบแบบไขว้ข้าม เป็นวิธีการตรวจสอบความผิดพลาดของโมเดลในการจำแนกอารมณ์ข้อความ บนข้อมูลชุดฝึกฝน เริ่มจากการสุ่มตัวอย่างโดยการแบ่งชุดข้อมุลออกเป็น K ส่วนเท่าๆ กัน หนึ่งส่วนสำหรับชุดข้อมูลทดสอบ ส่วนที่เหลือ K-1 ส่วน สำหรับชุดข้อมุลฝึกฝนแล้วทำการสร้างโมเดล และ ทดสอบการจำแนกอารมณ์ข้อความ หลังจากนั้นทำการสับเปลี่ยนชุดข้อมูลทดสอบไปจนกระทั่งครบ K รอบ

ในการทดลองนี้ ใช้ชุดข้อมูลสำหรับฝึกฝน จำนวน 3,051 ข้อความ นำมาสร้างโมเดลสำหรับฝึกฝน โดยใช้วิธีการหาฟังก์ชันความหนาแน่นของความน่าจะเป็น สำหรับการกระจายตัวเกาส์เซียนแบบปกติ (Probability Density Function for Normal Gaussian Distribution) และใช้อัลกอริทึมของนาอีฟเบย์ (Naïve Bayes algorithm) ในการจำแนกอารมณ์ของข้อความ สำหรับการทดสอบแบบไขว้ข้าม เลือกค่า K ที่ใช้ในการทดลอง คือ 4, 7 และ 10 ตามลำดับ ดังแสดงตารางที่ 4.1.1

ตารางที่ 4.1.1 การทดสอบแบบไขว้ข้าม
เปรียบเทียบระหว่างการให้น้ำหนักคำแบบ $tf.\chi 2$ และ $tf.\delta\text{-}\chi 2$

จำนวนชุดข้อมูลที่ถูกแบ่ง (K-Folds)	การให้น้ำหนักของคำ	
	$tf.\chi 2$	$tf.\delta\text{-}\chi 2$
4	63.69%	74.31%
7	63.27%	76.58%
10	64.33%	75.31%

ส่วนต่อไป เป็นการทดสอบหาค่าความแตกต่างเชิงนัยสำคัญ (Significant) สำหรับเปรียบเทียบวิธีการเรียนรู้ทั้ง 2 วิธีการ ใช้สำหรับการทดสอบข้อมูลที่มีการแจกแจงแบบปกติ (Normal Distribution) ผ่านการทดสอบแบบ Paired t-test [34] เริ่มจากการแสดงค่าความแม่นยำ และ ค่าความผิดพลาด แต่ละชุดข้อมูลทดสอบ ของการทดสอบแบบไขว้ข้าม ดังแสดงตาราง 4.1.2

ตารางที่ 4.1.2 ตารางแสดงค่าความแม่นยำ และ ค่าความผิดพลาด แต่ละชุดข้อมูลทดสอบ ของการทดสอบแบบไขว้ข้าม เปรียบเทียบระหว่างการให้น้ำหนักคำแบบ $tf.\chi 2$ และ $tf.\delta\text{-}\chi 2$

ข้อมูลทดสอบชุดที่	การให้น้ำหนักของคำ				ค่าความแตกต่างของค่าความผิดพลาด ($Error_{Ti}(h_A)$- $Error_{Ti}(h_B)$, δ_i)	ค่า $(\delta_i - \bar{\delta})^2$
	$tf.\chi 2$		$tf.\delta\text{-}\chi 2$			
	ค่าความแม่นยำ	ค่าความผิดพลาด ($Error_{Ti}(h_A)$)	ค่าความแม่นยำ	ค่าความผิดพลาด ($Error_{Ti}(h_B)$)		
1	0.6608	0.3392	0.7935	0.2065	0.1327	0.0002
2	0.6873	0.3127	0.8024	0.1976	0.1151	0.0000
3	0.7375	0.2625	0.8024	0.1976	0.0649	0.0028
4	0.6696	0.3304	0.7581	0.2419	0.0885	0.0009
5	0.5959	0.4041	0.6903	0.3097	0.0944	0.0005
6	0.6991	0.3009	0.8230	0.1770	0.1239	0.0000
7	0.5634	0.4366	0.6726	0.3274	0.1092	0.0001
8	0.5841	0.4159	0.7286	0.2714	0.1445	0.0007
9	0.6106	0.3894	0.7198	0.2802	0.1092	0.0001
10	0.4956	0.5044	0.6903	0.3097	0.1947	0.0059
					$\bar{\delta}$ = 0.1177	

$$S_{\bar{\delta}} \equiv \sqrt{\frac{1}{k(k-1)}\sum_{i=1}^{k}(\delta_i - \bar{\delta})^2} \quad (1)$$

โดยที่ $S_{\bar{\delta}}$ คือ ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)
k คือ จำนวนชุดข้อมูลทดสอบ
δ_i คือ ค่าความแตกต่างของค่าความผิดพลาด ในแต่ละรอบที่ i
$\bar{\delta}$ คือ ค่าเฉลี่ยของค่าความแตกต่างของค่าความผิดพลาด

จากค่า $(\delta_i - \bar{\delta})^2$ ในตาราง 4.1.2 และ ค่า k เท่ากับ 10 แทนค่าในสมการ (1) จะได้ค่า $S_{\bar{\delta}}$ เท่ากับ 0.0112

ตารางที่ 4.1.3 ตารางค่า Z_N สำหรับช่วงความเชื่อมั่น $N\%$
สำหรับข้อมูลที่มีการกระจายตัวความน่าจะเป็นแบบปกติ (Normal Probability Distribution)

ระดับความเชื่อมั่น $N\%$	50%	68%	80%	90%	95%	98%	99%
ค่า Z_N	0.67	1.00	1.28	1.64	1.96	2.33	2.58

จากตาราง 4.1.3 ช่วงความเชื่อมั่น (Confidence interval) ที่ 95% ให้ค่า Z_N หรือ $t_{N,k-1}$ เท่ากับ 1.96 แทนค่าในสมการ (2)

$$\bar{\delta} \pm t_{N,k-1} S_{\bar{\delta}} \quad (2)$$

จากค่า $\bar{\delta}$ เท่ากับ 0.1177, ค่า $S_{\bar{\delta}}$ เท่ากับ 0.0112 และ $t_{N,k-1}$ เท่ากับ 1.96 แทนค่าในสมการ (2) จะได้ว่าค่าเฉลี่ยของค่าความแตกต่างของค่าความผิดพลาด ($\bar{\delta}$) เท่ากับ 0.1177 $\pm$ 0.0219 บนช่วงความเชื่อมั่น 95% กล่าวอีกอย่างหนึ่ง คือ มีค่านัยสำคัญ (Significant) เท่ากับ 1 - 0.95 หรืออยู่ที่ 0.05

4.2. การทดสอบความแม่นยำ (Accuracy) บนแต่ละวิธีการให้น้ำหนักของคำ

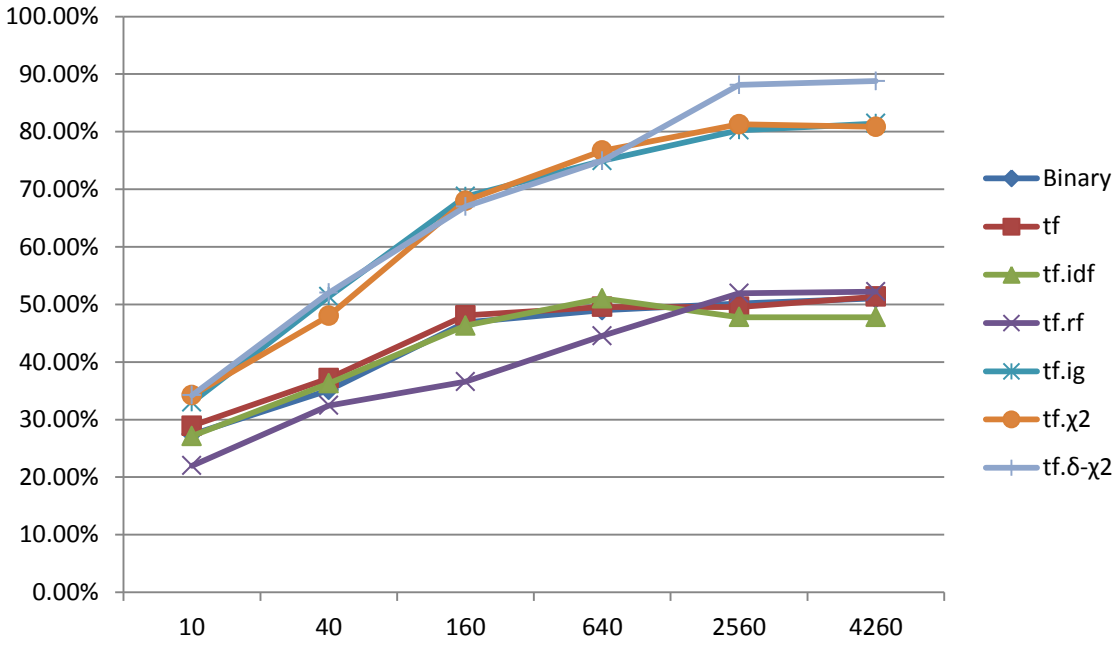
การทดสอบความแม่นยำ เป็นวิธีการตรวจสอบความผิดพลาดของโมเดลในการจำแนกอารมณ์ข้อความ บนข้อมูลชุดทดสอบ ค่าความแม่นยำ ($Accuracy$) คือ จำนวนข้อความที่ระบบจำแนกข้อความได้ถูกต้อง เทียบกับ จำนวนข้อความทั้งหมดบนข้อมูลชุดทดสอบ ดังสมการที่ 3

$$Accuracy(\%) = \frac{\#\ Correctly\ Classified\ Document}{\#\ Total\ Test\ Document} \times 100 \quad (3)$$

ในการทดลองนี้ ใช้ชุดข้อมูลสำหรับฝึกฝน จำนวน 3,051 ข้อความ นำมาสร้างโมเดลสำหรับฝึกฝน โดยใช้วิธีการหาฟังก์ชันความหนาแน่นของความน่าจะเป็น สำหรับการกระจายตัวเกาส์เซียนแบบปกติ (Probability Density Function for Normal Gaussian Distribution) [10] และใช้อัลกอริทึมของนาอีฟเบย์ (Naïve Bayes algorithm) [10] ในการจำแนกอารมณ์ของชุดข้อมูลสำหรับทดสอบ จำนวน 339 ข้อความ ซึ่งจำนวนดังกล่าวเป็นข้อความคนละชุดกับข้อมูลสำหรับฝึกฝน

ตารางที่ 4.2.1 การทดสอบความแม่นยำ บนข้อมูลชุดทดสอบ
เปรียบเทียบการให้น้ำหนักคำ $Binary$, tf, $tf.idf$, $tf.rf$, $tf.ig$, $tf.\chi 2$ และ $tf.\delta\text{-}\chi 2$

การให้น้ำหนักของคำ	จำนวนคำที่ถูกคัดเลือก จากอันดับคะแนนมากไปน้อย					
	10	40	160	640	2,560	4,260
$Binary$	27.44%	35.10%	46.90%	48.97%	50.15%	51.03%
tf	28.91%	37.17%	48.08%	49.56%	49.56%	51.33%
$tf.idf$	27.14%	36.28%	46.31%	51.03%	47.79%	47.79%
$tf.rf$	22.00%	32.45%	36.58%	44.54%	51.92%	52.21%
$tf.ig$	33.00%	51.33%	68.73%	74.93%	80.24%	81.42%
$tf.\chi 2$	34.21%	48.00%	68.00%	76.70%	81.29%	80.83%
$tf.\delta\text{-}\chi 2$	34.21%	52.00%	67.00%	74.93%	88.13%	88.79%



ภาพที่ 4.1 แสดงความสัมพันธ์ระหว่างความแม่นยำ กับ จำนวนคำที่มีการเรียงคะแนน
เปรียบเทียบการให้น้ำหนักคำ $Binary$, tf, $tf.idf$, $tf.rf$, $tf.ig$, $tf.\chi 2$ และ $tf.\delta\text{-}\chi 2$

จากตารางที่ 4.2.1 แสดงผลทดสอบความแม่นยำในการจำแนกข้อความบนการให้น้ำหนักคำแบบต่างๆ สามารถแบ่งออกเป็นสองกลุ่มหลัก คือ กลุ่มการให้น้ำหนักคำแบบไม่มีผู้สอน (Unsupervised term weighting) ได้แก่ $Binary$, tf, $tf.idf$ กับ กลุ่มการให้น้ำหนักคำแบบมีผู้สอน (Supervised term weighting) ได้แก่ $tf.rf$, $tf.ig$, $tf.\chi 2$, $tf.\delta\text{-}\chi 2$ จากการทดลองพบว่าทั้งสองกลุ่มเมื่อจำนวนคำเพิ่มขึ้นทำให้ความแม่นยำเพิ่มขึ้นตามด้วย แต่เมื่อจำนวนคำที่เพิ่มขึ้นเข้าใกล้จำนวนคำทั้งหมดเท่าไร จะทำให้อัตราการเพิ่มความแม่นยำจะลดลงเท่านั้น เมื่อเปรียบเทียบกับจำนวนที่น้อยกว่า แล้วโดยรวมกลุ่มการให้น้ำหนักคำแบบมีผู้สอน จะให้ผลความแม่นยำที่ดีกว่ากลุ่มการให้น้ำหนักคำแบบไม่มีผู้สอน ส่วนความแม่นยำของจำนวน 2,560 กับ 4,260 คำ จะพบว่ามีค่าความแม่นยำที่ใกล้เคียงกัน นั้นแสดงให้เห็นว่า จำนวน 2,560 คำที่มีการเรียงคะแนนของคำ มีประสิทธิภาพการจำแนกอารมณ์บนข้อความ เทียบเท่ากับจำนวนคำทั้งหมด 4,260 คำ

จากผลการทดลองที่แล้ว ความแม่นยำในการจำแนกข้อความจากวิธีการให้น้ำหนักแบบเดลต้า-ไคสแควร์ อยู่ที่ 88.79% บนจำนวนคำ 4,260 คำ ในส่วนต่อไป จะเป็นการวิเคราะห์ความผิดพลาดในการจำแนกอารมณ์จากข้อความ (Error Analysis) บนวิธีการให้น้ำหนักแบบเดลต้า-ไคสแควร์ อยู่ที่ 11.21% ซึ่งเป็นข้อความจำนวน 38 ข้อความ จากชุดข้อความทดสอบจำนวน 339 ข้อความ สามารถจำแนกสาเหตุความผิดพลาดในการจำแนกอารมณ์จากข้อความ ดังตาราง 4.2.2

ตารางที่ 4.2.2 แสดงการวิเคราะห์ความผิดพลาด (Error Analysis)
ในการจำแนกอารมณ์บนข้อความ

จำนวนข้อความ	สาเหตุความผิดพลาด	ตัวอย่างข้อความ
4	มีคำที่ไม่ปรากฏบนชุดฝึกฝน หรือ คำที่ปรากฏไม่เกิน 10 ข้อความ	ม่าย| |กรี๊ด| |ทำร้าย|จิตใจ|มาก|
16	ไม่มีคำที่มีลักษณะอารมณ์เด่น ที่สอดคล้องกับอารมณ์ของข้อความ	รู้|ว่า|ประเทศ|ไทย|ป่า|เหลือ|น้อย|สุด|แล้ว|
18	มีคำที่มีลักษณะอารมณ์เด่น ที่สอดคล้องกับอารมณ์ของข้อความ แต่มีคำที่มีลักษณะอารมณ์อื่นที่เด่นกว่า	เริ่ม|เล่น|ไม่|สนุก|เลย||เซ็ง|มาก|

จากตารางที่ 4.2.2 พบว่าสาเหตุความผิดพลาดในการจำแนกอารมณ์บนข้อความแบ่งออกเป็น 3 ข้อ ได้แก่ 1) คำที่ไม่เคยเกิดขึ้น หรือ เกิดขึ้นน้อยบนข้อมูลชุดฝึกฝน 2) ไม่มีคำที่มีลักษณะอารมณ์เด่น และ 3) มีคำที่มีลักษณะอารมณ์อื่นที่เด่นกว่า ต่อไปเป็นการวิเคราะห์สาเหตุแต่ละข้อ จากตัวอย่างข้อความ โดยการแจกแจงจำนวนความถี่ของคำ ในแต่ละอารมณ์ บนตารางที่ 4.2.2.1 - 4.2.2.3

ตารางที่ 4.2.2.1 แสดงการแจกแจงจำนวนความถี่ของคำ ในแต่ละอารมณ์
บนข้อความ “ม่าย| |กรี๊ด| |ทำร้าย|จิตใจ|มาก|”

		อารมณ์						จำนวนข้อความ
		ร่าเริง	เสียใจ	โกรธ	กลัว	รังเกียจ	ประหลาดใจ	
ตัวอย่างคำ	ม่าย	0	7	0	3	0	0	10
	กรี๊ด	0	0	0	3	0	0	3
	ทำร้าย	0	3	0	0	0	0	3
	จิตใจ	0	3	0	0	0	0	3
	มาก	93	40	12	126	16	10	297

ต่อไปจะเป็นการวิเคราะห์สาเหตุแต่ละข้อ สาเหตุแรก มีคำที่ไม่เคยเกิดขึ้น หรือ เกิดขึ้นน้อย บนข้อมูลชุดฝึกฝน จากการทดลองได้ทำการแยกชุดข้อมูล ออกเป็นชุดข้อมูลฝึกฝนจำนวน 3,051 ข้อความ กับ ชุดข้อมูลทดสอบจำนวน 339 ข้อความ ซึ่งทั้งสองชุดข้อมูลแยกออกจากกัน จึงมีความเป็นไปได้ว่า คำที่ไม่เคยปรากฏบนชุดฝึกฝน หรือ เกิดขึ้นน้อย อาจจะปรากฏบนชุดข้อมูลทดสอบ ทำให้ระบบไม่สามารถวิเคราะห์คำดังกล่าว ได้ว่าคำนั้นมีลักษณะอารมณ์เด่นหรือไม่ จึงมีโอกาสที่ระบบจะทำนายอารมณ์บนข้อความออกมาผิดพลาด ตัวอย่างข้อความ “เสียใจ” เช่น **ม่าย| |กรี๊ด| |ทำร้าย| จิตใจ|มาก|** แต่ระบบทำนายอารมณ์บนข้อความ คือ “กลัว” และ ทำการแจกแจงจำนวนความถี่ของคำ ในแต่ละอารมณ์ ดังตาราง 4.2.2.1 พบว่า มีคำที่ไม่ปรากฏบนชุดฝึกฝน หรือ คำที่ปรากฏไม่เกิน 10 ข้อความ ได้แก่ “**ม่าย**”, “**กรี๊ด**”, “**ทำร้าย**” และ “**จิตใจ**” แต่จะมีคำว่า “**มาก**” ที่ปรากฏขึ้นจำนวน 297 ข้อความ และ เป็นคำที่มีลักษณะอารมณ์เด่น “กลัว” ดังนั้นระบบจึงทำนายอารมณ์บนข้อความผิดพลาด เป็นอารมณ์ “กลัว” แทนที่จะเป็นอารมณ์ที่ถูกต้อง คือ อารมณ์ “เสียใจ” (ซึ่งได้จากการโหวตของอาสาสมัคร)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.2.2.2 แสดงการแจกแจงจำนวนความถี่ของคำ ในแต่ละอารมณ์
บนข้อความ “รู้|ว่า|ประเทศ|ไทย|ป่า|เหลือ|น้อย|สุด|แล้ว|”

		อารมณ์						จำนวนข้อความ
		ร่าเริง	เสียใจ	โกรธ	กลัว	รังเกียจ	ประหลาดใจ	
ข้อความ	รู้	5	3	5	6	3	19	41
	ว่า	6	8	11	8	19	22	74
	ประเทศ	0	0	5	0	2	0	7
	ไทย	0	6	6	2	3	4	21
	ป่า	0	0	1	2	0	0	3
	เหลือ	0	5	2	3	1	1	12
	น้อย	0	3	0	0	2	0	5
	สุด	17	13	2	**29**	0	0	61
	แล้ว	42	47	19	**70**	19	24	221

สาเหตุที่สอง ไม่มีคำที่มีลักษณะอารมณ์เด่น เนื่องจากระบบวิเคราะห์การจำแนกอารมณ์บนข้อความ จากความน่าจะเป็นที่ปรากฏร่วมกันระหว่างคำ กับ อารมณ์ เพื่อหาว่าอารมณ์ใดมีลักษณะการปรากฏอย่างเด่นชัดบนคำนั้น ดังนั้นหากไม่มีคำที่มีลักษณะอารมณ์เด่นปรากฏขึ้น ระบบมีโอกาสจะทำนายอารมณ์บนข้อความออกมาผิดพลาด ตัวอย่างข้อความ “เสียใจ” เช่น **รู้|ว่า|ประเทศ|ไทย|ป่า|เหลือ|น้อย|สุด|แล้ว|** แต่ระบบทำนายอารมณ์บนข้อความ คือ “กลัว” และ ทำการแจกแจงจำนวนความถี่ของคำ ในแต่ละอารมณ์ ดังตาราง 4.2.2.2 พบว่า ไม่มีคำที่มีลักษณะอารมณ์เด่น ที่สอดคล้องกับอารมณ์เสียใจบนข้อความดังกล่าว แต่มีคำว่า “**สุด**” กับ “**แล้ว**” เป็นคำที่มีลักษณะอารมณ์เด่น “กลัว” ดังนั้นระบบจึงทำนายอารมณ์บนข้อความผิดพลาด เป็นอารมณ์ “กลัว” แทนที่จะเป็นอารมณ์ที่ถูกต้อง คือ อารมณ์ “เสียใจ” (ซึ่งได้จากการโหวตของอาสาสมัคร)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.2.2.3 แสดงการแจกแจงจำนวนความถี่ของคำ ในแต่ละอารมณ์
บนข้อความ “เริ่ม|เล่น|ไม่|สนุก|เลย||เซ็ง|มาก|”

		อารมณ์						จำนวนข้อความ
		ร่าเริง	เสียใจ	โกรธ	กลัว	รังเกียจ	ประหลาดใจ	
ตัวอย่างคำ	เริ่ม	0	2	0	3	1	9	15
	เล่น	12	6	1	6	2	2	29
	ไม่	11	**185**	35	108	71	52	462
	สนุก	27	3	0	0	0	0	30
	เลย	69	53	29	107	27	24	309
	เซ็ง	0	3	0	0	**7**	0	10
	มาก	93	40	12	126	16	10	297

สาเหตุที่สาม ถึงแม้จะมีคำที่มีลักษณะอารมณ์เด่น ที่สอดคล้องกับอารมณ์ของข้อความ แต่มีคำที่มีลักษณะอารมณ์อื่นที่เด่นกว่า ระบบวิเคราะห์การจำแนกอารมณ์บนข้อความ จากความน่าจะเป็นที่ปรากฏร่วมกันระหว่างคำกับอารมณ์ ของทุกคำที่ปรากฏบนข้อความ ดังนั้นระบบมีโอกาสจะทำนายอารมณ์บนข้อความออกมาผิดพลาด จากคำที่มีลักษณะอารมณ์อื่นที่เด่นกว่า ตัวอย่างข้อความ “รังเกียจ” เช่น **เริ่ม|เล่น|ไม่|สนุก|เลย||เซ็ง|มาก|** แต่ระบบทำนายอารมณ์บนข้อความ คือ “เสียใจ” และ ทำการแจกแจงจำนวนความถี่ของคำ ในแต่ละอารมณ์ ดังตาราง 4.2.2.3 พบว่า มีคำที่มีลักษณะอารมณ์เด่น ที่สอดคล้องกับอารมณ์รังเกียจของข้อความคือ คำว่า “**เซ็ง**” แต่มีคำที่มีลักษณะอารมณ์อื่นที่เด่นกว่าคือ คำว่า “**ไม่**” เป็นคำที่มีลักษณะอารมณ์เด่น “เสียใจ” ดังนั้นระบบจึงทำนายอารมณ์บนข้อความผิดพลาด เป็นอารมณ์ “เสียใจ” แทนที่จะเป็นอารมณ์ที่ถูกต้อง คือ อารมณ์ “รังเกียจ” (ซึ่งได้จากการโหวตของอาสาสมัคร)

ส่วนต่อไป จะเป็นการวิเคราะห์การทดสอบความแม่นยำ ของคำที่ไม่ได้ถูกคัดเลือกจากอันดับคำ 2,561 ถึง 4,260 เปรียบเทียบการให้น้ำหนักคำระหว่าง $tf.\chi 2$ และ $tf.\delta\text{-}\chi 2$

ตารางที่ 4.3 การทดสอบความแม่นยำ ของคำที่ไม่ได้ถูกคัดเลือก
เปรียบเทียบการให้น้ำหนักคำระหว่าง $tf.\chi 2$ และ $tf.\delta$-$\chi 2$

การให้น้ำหนักของคำ	ช่วงอันดับคะแนนของคำ ที่ถูกคัดเลือก		
	อันดับ 1-2,560 (คำที่ถูกคัดเลือก)	อันดับ 2,561-4,260 (คำที่ไม่ได้ถูกคัดเลือก)	อันดับ 1-4,260 (คำทั้งหมด)
$tf.\chi 2$	81.12%	17.40%	80.83%
$tf.\delta$-$\chi 2$	88.79%	16.52%	88.79%

ตารางที่ 4.4 จำนวนรูปแบบคำ และ จำนวนคำ ที่อยู่บนชุดข้อความทดสอบ
แต่ไม่ได้อยู่บนชุดข้อความฝึกฝน

		ช่วงอันดับคะแนนของคำ ที่ถูกคัดเลือก	
		อันดับ 1-2,560	อันดับ 1-4,260
การให้น้ำหนักของคำ $tf.\chi 2$	จำนวนรูปแบบคำ (ไม่นับคำซ้ำ)	583	378
	จำนวนคำ (นับคำซ้ำ)	655	400
การให้น้ำหนักของคำ $tf.\delta$-$\chi 2$	จำนวนรูปแบบคำ (ไม่นับคำซ้ำ)	613	378
	จำนวนคำ (นับคำซ้ำ)	694	400
	จำนวนรูปแบบคำทั้งหมด (ไม่นับคำซ้ำ)	1,338	
	จำนวนคำทั้งหมด (นับคำซ้ำ)	4,750	

ในส่วนอันดับรูปแบบของคำที่เหลือ อันดับ 2,561-4,260 จะวิเคราะห์ความแม่นยำ ใน ตาราง 4.3 ซึ่งให้ผลความแม่นยำบนวิธีการให้น้ำหนักของคำแบบ $tf.\chi 2$ กับ $tf.\delta$-$\chi 2$ คือ 17.40% กับ 16.52% ตามลำดับ ความแม่นยำดังกล่าว เทียบเคียงได้กับความน่าจะเป็นในการสุ่มอารมณ์บนข้อความจาก 6 อารมณ์ คือ 16.67% (100% หารด้วย 6) แสดงให้เห็นว่า อันดับคำที่ 2,561-4,260 มีผลต่อการจำแนกอารมณ์บนข้อความ ใกล้เคียงกับการสุ่มอารมณ์บนข้อความจาก 6 อารมณ์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

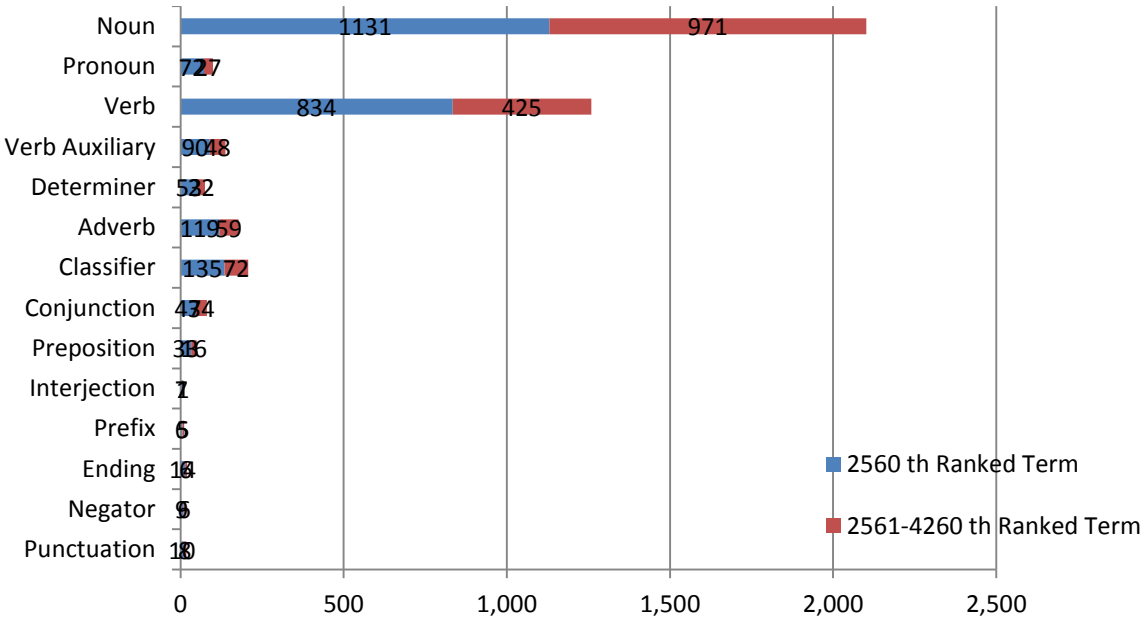
ในส่วนตารางที่ 4.4 แสดงจำนวนรูปแบบคำ และ จำนวนคำ ที่อยู่บนชุดข้อความทดสอบ แต่ไม่ได้อยู่บนชุดข้อความฝึกฝน จากตารางแสดงให้เห็นว่า ยังคงมีจำนวนรูปแบบคำที่ยังไม่เคยได้รับการฝึกฝนข้อมูลจากระบบอยู่

ในส่วนต่อไป จะเป็นการวิเคราะห์การแจกแจงจำนวนคำบนแต่ละหน้าที่ของคำ (Part of Speech, POS) [18] บนวิธีการคัดเลือกคำแบบไคสแควร์ กับ วิธีการคัดเลือกคำแบบเดลต้า-ไคสแควร์ ในงานวิจัยนี้ได้เลือกใช้วิธีการตัดคำภาษาไทย และ ระบุหน้าที่ของคำผ่านโปรแกรม SWATH [33] โดยเลือกใช้อัลกอริทึมไบแกรม (Bigram) เป็นการประมาณค่าความน่าจะเป็นของชุดอักขระที่เกิดขึ้นร่วมกันว่ามีค่าเท่ากับ ผลคูณของความน่าจะเป็นที่จะพบอักขระทีละ 2 ตัว ติดกันในชุดอักขระนั้น

ตารางที่ 4.5 แสดงการแจกแจงจำนวนคำบนแต่ละหน้าที่ของคำ (Part of Speech, POS) บนวิธีการคัดเลือกคำแบบไคสแควร์ ($\chi 2$)

หน้าที่ของคำ	จำนวนคำที่มีการเรียงคะแนน	
(Part of Speech, POS)	2,560	4,260
Noun	1,131	2,102
Pronoun	72	99
Verb	834	1,259
Verb Auxiliary	90	138
Determiner	53	75
Adverb	119	178
Classifier	135	207
Conjunction	47	81
Preposition	33	49
Interjection	7	8
Prefix	6	11
Ending	14	20
Negator	9	15
Punctuation	10	18

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



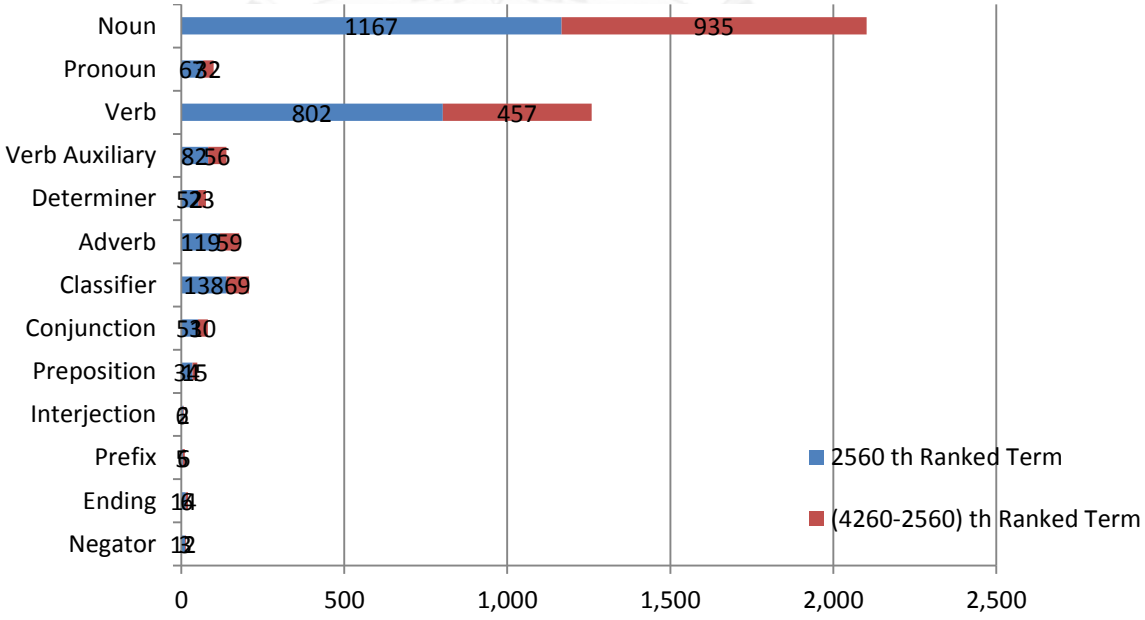
ภาพที่ 4.2 แสดงการแจกแจงจำนวนคำบนแต่ละหน้าที่ของคำ (Part of Speech, POS) บนวิธีการคัดเลือกคำแบบไคสแควร์ ($\chi 2$)

จากตารางที่ 4.5 แสดงการแจกแจงจำนวนคำบนแต่ละหน้าที่ของคำ (Part of Speech, POS) บนวิธีการคัดเลือกคำแบบไคสแควร์ ($\chi 2$) แนวคิดการจำแนกหน้าที่ของคำใช้ตามแนวคิดของ ORCHID [18] ซึ่งแบ่งออกเป็น 14 กลุ่ม คือ Noun, Pronoun, Verb, Verb Auxilliary, Determiner, Adverb, Classifier, Conjunction, Preposition, Interjection, Prefix, Ending, Negator และ Punctuation จากการทดลองพบว่า คำที่มีการเรียงคะแนนจำนวน 2,560 คำ มีการกระจายตัวอยู่ในทุกหน้าที่ของคำ โดยที่จำนวนหน้าที่ของคำที่ถูกคัดเลือกมามากที่สุด คือ คำนาม (Noun) กับ คำกริยา (Verb) ในส่วนวิธีการคัดเลือกคำแบบเดลต้า-ไคสแควร์ (δ-$\chi 2$) จะให้ผลการการแจกแจงที่ใกล้เคียงดังแสดงตารางที่ 4.6

Chulalongkorn University

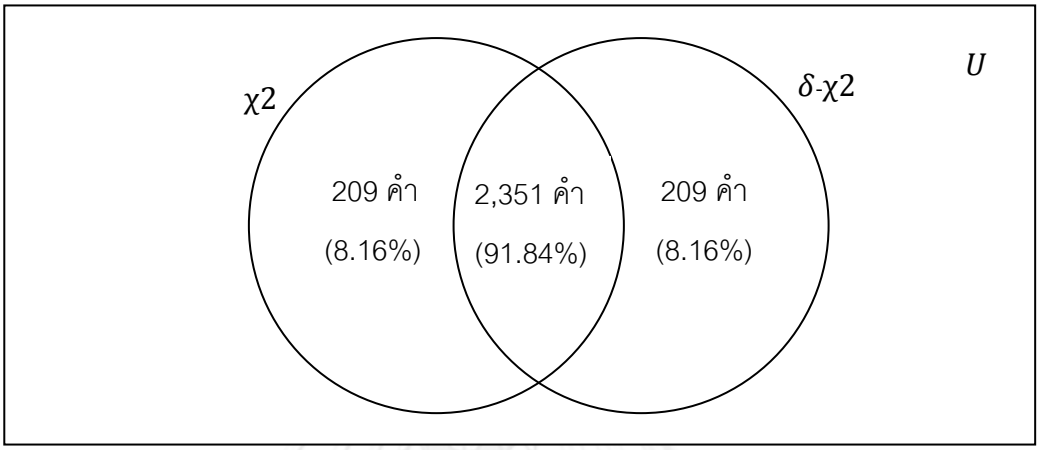
ตารางที่ 4.6 แสดงการแจกแจงจำนวนคำของแต่ละหน้าที่ของคำ (Part of Speech, POS) บนวิธีการคัดเลือกคำแบบเดลต้า-ไคสแควร์ (δ-$\chi 2$)

หน้าที่ของคำ (Part of Speech, POS)	จำนวนคำที่มีการเรียงคะแนน	
	2,560	4,260
Noun	1,167	2,102
Pronoun	67	99
Verb	802	1,259
Verb Auxiliary	82	138
Determiner	52	75
Adverb	119	178
Classifier	138	207
Conjunction	51	81
Preposition	34	49
Interjection	6	8
Prefix	5	11
Ending	14	20
Negator	12	15
Punctuation	11	18



ภาพที่ 4.3 แสดงการแจกแจงจำนวนคำของแต่ละหน้าที่ของคำ (Part of Speech, POS) บนวิธีการคัดเลือกคำแบบเดลต้า-ไคสแควร์ (δ-$\chi 2$)

ในส่วนต่อไป จะวิเคราะห์ความแตกต่างของคำระหว่างวิธีการคัดเลือกคำแบบไคสแควร์ กับวิธีการคัดเลือกคำแบบเดลต้า-ไคสแควร์ บนจำนวนรูปแบบคำ 2,560 คำ ที่มีการเรียงคะแนนมาแล้ว โดยพิจารณาการแจกแจงจำนวนความถี่ของคำที่เกิดขึ้น บนแต่ละอารมณ์ ดังแสดงตารางที่ 4.7 และ 4.8



ภาพที่ 4.4 แสดงความแตกต่างของรูปแบบคำ อันดับที่ 1-2,560
ระหว่างการคัดเลือกคำแบบ $\chi 2$ กับ δ-$\chi 2$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.7 แสดงการแจกแจงจำนวนความถี่ของคำ ในแต่ละอารมณ์
บนวิธีการคัดเลือกคำแบบไคสแควร์ ($\chi 2$)

อันดับของคำ	รูปแบบของคำ	อารมณ์บนข้อความ					
		ร่าเริง	เสียใจ	โกรธ	กลัว	รังเกียจ	ประหลาดใจ
1	น่ากลัว@ADVN	0	0	3	332	1	0
2	หลอน@VSTA	0	0	0	130	0	0
7	สงสาร@VACT	0	57	0	0	0	0
10	ผี@NCMN	0	0	2	57	5	2
13	ไอ้@PPRS	0	0	59	3	11	3
153	จะ@XVBM	20	19	30	43	35	24
156	เป็น@VSTA	16	23	6	11	13	21
157	เพราะ@JSBR	8	21	1	2	2	2
400	ค่ะ@EAFF	12	10	1	23	1	1
458	มี@VSTA	33	30	10	26	26	29
492	ที่@PREL	29	29	16	23	21	6
531	คน@CNIT	36	36	22	30	20	16
2,483	แล้ว@XVAE	43	44	23	50	19	26
...	...	...	...	...	...	...	...

คำที่กระจายตัวอยู่ในทุกอารมณ์

เส้นกำกับอันดับของคำที่ 2,560

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.8 แสดงการแจกแจงจำนวนความถี่ของคำ ในแต่ละอารมณ์
บนวิธีการคัดเลือกคำแบบเดลต้า-ไคสแควร์ (δ-$\chi 2$)

อันดับของคำ	รูปแบบของคำ	อารมณ์บนข้อความ					
		ร่าเริง	เสียใจ	โกรธ	กลัว	รังเกียจ	ประหลาดใจ
1	น่ากลัว@ADVN	0	0	3	332	1	0
2	หลอน@VSTA	0	0	0	130	0	0
7	สงสาร@VACT	0	57	0	0	0	0
9	ผี@NCMN	0	0	2	57	5	2
14	ไอ้@PPRS	0	0	59	3	11	3
174	เพราะ@JSBR	8	21	1	2	2	2
430	ค่ะ@EAFF	12	10	1	23	1	1
1,636	จะ@XVBM	20	19	30	43	35	24
2,561	มี@VSTA	33	30	10	26	26	29
2,851	ที่@PREL	29	29	16	23	21	6
2,967	แล้ว@XVAE	43	44	23	50	19	26
2,972	เป็น@VSTA	16	23	6	11	13	21
3,017	คน@CNIT	36	36	22	30	20	16
...	...	...	...	...	...	...	...

เส้นกำกับ
อันดับของคำที่
2,560

คำที่กระจายตัว
อยู่ในทุกอารมณ์

จากตารางที่ 4.7-4.8 แสดงการแจกแจงจำนวนความถี่ของคำ ในแต่ละอารมณ์ บนวิธีการคัดเลือกคำแบบ ไคสแควร์ ($\chi 2$) และ บนวิธีการคัดเลือกคำแบบเดลต้า-ไคสแควร์ (δ-$\chi 2$) พบว่าคำที่มีความแตกต่างระหว่างสองวิธี (ที่ทำลูกศรชี้) ซึ่งคำเหล่านั้นมีลักษณะการกระจายตัวอยู่ในทุกอารมณ์อย่างใกล้เคียงกัน บนวิธีการคัดเลือกคำแบบไคสแควร์ แต่ขณะที่วิธีการคัดเลือกคำแบบเดลต้า-ไคสแควร์ จะพบคำเหล่านั้นน้อยกว่า เนื่องจากวิธีการคัดเลือกคำแบบเดลต้า-ไคสแควร์ จะทำการพิจารณาคู่ของค่า $n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k)$ กับ $n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k})$ ซึ่งถ้าค่า $n(T_j, \overline{E_k}) \times n(\overline{T_j}, E_k)$ มีค่ามากกว่า $n(T_j, E_k) \times n(\overline{T_j}, \overline{E_k})$ จะทำให้มีโอกาสเกิดคำที่มีลักษณะ Multi-Emotion มากขึ้น ดังนั้นเมื่อเกิดกรณีดังกล่าว จะเข้ากรณี δ เป็นศูนย์ ทำให้คะแนนของคู่คำกับอารมณ์นั้น $Score_{\delta-\chi 2}(T_j, E_k)$ เป็นศูนย์ตาม ส่งผลให้การคิดค่าเฉลี่ย $Score_{average}(T_j)$ มีค่าลดลงตามด้วย เมื่อมีการเรียงคะแนนทำให้อันดับคะแนนของคำนั้นตกลงด้วย ให้คงเหลือชุด

ของคำที่มีคุณลักษณะที่เกิดขึ้นในอารมณ์เฉพาะมากขึ้น ดังแสดงตารางที่ 4.8 ส่วนส่วนต่อไปจะวิเคราะห์ความแตกต่างของการให้น้ำหนักของคำแบบไคสแควร์ ($tf.\chi 2$) กับ แบบเดลต้า-ไคสแควร์ ($tf.\delta$-$\chi 2$)

ตารางที่ 4.9 แสดงการให้น้ำหนักของคำแบบ $tf.\chi 2$

	รูปแบบของคำ							
	ไป	มาก	อะไร	มา	มี	ว่า	ที่	...
ข้อความที่ 1	0	0.053	0	0	0.960	0	0.003	...
ข้อความที่ 2	2.801	0	0	0	0.018	0	0	...
ข้อความที่ 3	24.069	0	0	0	0	0.006	3.433	...
ข้อความที่ 4	0	0	17.309	0	0.277	2.654	0.008	...
ข้อความที่ 5	0	0	0	0.024	0	0	0	...

ตารางที่ 4.10 แสดงการให้น้ำหนักของคำแบบ $tf.\delta$-$\chi 2$

	รูปแบบของคำ							
	ไป	มาก	อะไร	มา	มี	ว่า	ที่	...
ข้อความที่ 1	0	0	0	0	0.960	0	0	...
ข้อความที่ 2	0	0	0	0	0	0	0	...
ข้อความที่ 3	0	0	0	0	0	0	3.433	...
ข้อความที่ 4	0	0	17.309	0	0.277	2.654	0	...
ข้อความที่ 5	0	0	0	0	0	0	0	...

ตารางที่ 4.11 แสดงตัวอย่างข้อความของตารางที่ 4.7 กับ 4.8

ข้อความที่	ข้อความ
1	ใช่| |เธอ|เป็น|คน|บอก|รัก|ฉัน|ทำ|ให้|ฉัน|ดีใจ|และ|มี|ความ|สุข|มาก|
2	โกง|คอร์รัปชั่น| |ประหาร|ชีวิต| |ตืด|คุก|ตลอด|ชีวิต| |ไม่|มี|อภัยโทษ|เด็ดขาด|
3	ตอน|ที่|กล้อง|ล้ม|อ่ะ|ค่ะ| |ว่า|ตั้ง|นาน|ไม่|ล้ม| |อ่ะ| |เห็น|เงา| |ขาว|แวบ|
4	|มี|ใคร|พอ|ทราบ|มั้ย|ครับ|ว่า|มัน|คือ|อะไร||
5	เอา|ไม้|ฟาด|หน้า|จอ|เมื่อ|เจอ|พวก|มือบอน| |มา|โพสต์|ตอน|จบ|แกลัง|คน|

จากตารางที่ 4.9-4.11 แสดงการให้น้ำหนักของคำเปรียบเทียบการให้น้ำหนักของคำระหว่าง $tf.\chi 2$ กับ $tf.\delta$-$\chi 2$ พบว่าความแตกต่างเป็นไปตามกรอบสีแดง จากวิธีการให้น้ำหนักของคำแบบ $tf.\chi 2$ จะทำการให้น้ำหนักทุกคำที่เกิดขึ้นบนข้อความ ขณะที่วิธีการให้น้ำหนักของคำแบบ $tf.\delta$-$\chi 2$ จะพิจารณาก่อนว่าคำดังกล่าวมีลักษณะการกระจายในทุก

อารมณ์หรือไม่ จากพารามิเตอร์เดลต้า (δ) ถ้าค่าเดลต้าเป็นศูนย์ น้ำหนักของคำนั้นจะมีค่าเท่ากับศูนย์ แต่ถ้า δ ไม่เป็นศูนย์ น้ำหนักของคำจะถูกคำนวณตามปกติ

ตารางที่ 4.12 ตารางแสดงความแม่นยำการจำแนกอารมณ์ บนข้อความระหว่างเจ้าของข้อความ และ ไม่ใช่เจ้าของข้อความ

การให้น้ำหนักของคำ	การเรียนรู้แบบมีผู้สอน โดย			
	การโหวตทั้งเจ้าของ และไม่ใช่เจ้าของข้อความ	เจ้าของข้อความ	ไม่ใช่เจ้าของข้อความ	การจับคู่ระหว่างเจ้าของ กับ ไม่ใช่เจ้าของข้อความ
$tf.\chi 2$	84.36%	81.12%	85.25%	80.83%
$tf.\delta$-$\chi 2$	91.15%	91.15%	89.97%	88.79%

จากตารางที่ 4.12 แสดงความแม่นยำการจำแนกอารมณ์ บนข้อความ ระหว่างเจ้าของข้อความ และ ไม่ใช่เจ้าของข้อความ พบว่า วิธีการให้น้ำหนักของคำแบบ $tf.\delta$-$\chi 2$ เพิ่มความแม่นยำ จากวิธีการให้น้ำหนักของคำแบบ $tf.\chi 2$ บนทุกรูปแบบการเรียนรู้แบบมีผู้สอน แต่บนวิธีการให้น้ำหนักของคำเดียวกัน การเรียนรู้แบบมีผู้สอนแบบวิธีการโหวตทั้งเจ้าของ และ ไม่ใช่เจ้าของข้อความ ยังคงให้ความแม่นยำมากที่สุด บนทุกรูปแบบการเรียนรู้แบบมีผู้สอน ขณะที่การเรียนรู้แบบมีผู้สอนแบบวิธีการจับคู่ระหว่างเจ้าของ กับ ไม่ใช่เจ้าของข้อความ ให้ความแม่นยำน้อยที่สุด อย่างไรก็ตาม วิธีการให้น้ำหนักของคำแบบ $tf.\delta$-$\chi 2$ เทียบกับวิธีการให้น้ำหนักของคำแบบ $tf.\chi 2$ ได้ทำการลดความแตกต่างของความแม่นยำระหว่างการเรียนรู้แบบมีผู้สอนทั้งสองวิธีการดังกล่าว โดยบนวิธีการให้น้ำหนักของคำแบบ $tf.\chi 2$ ให้ความแตกต่างของความแม่นยำทั้งสองวิธีการ อยู่ที่ 3.53% ในขณะที่บนวิธีการให้น้ำหนักของคำแบบ $tf.\delta$-$\chi 2$ ให้ความแตกต่างของความแม่นยำทั้งสองวิธีการ ลดลงอยู่ที่ 2.36%

ในส่วนต่อไป จะวิเคราะห์ผลการจำแนกอารมณ์บนข้อความ โดยการแจกแจงความแม่นยำของแต่ละอารมณ์ บนวิธีการคัดเลือกแบบไคสแควร์ และ เดลต้า-ไคสแควร์ ดังตาราง 4.13 - 4.14

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.13 ตารางแสดงการแจกแจงความแม่นยำของแต่ละอารมณ์
บนวิธีการคัดเลือกแบบไคสแควร์

		อารมณ์ที่ทำนาย (จากโปรแกรม)						ความแม่นยำ	ความแม่นยำเป็น %
		ร่าเริง / ดีใจ	เสียใจ	โกรธ	กลัว	รังเกียจ	ประหลาดใจ		
อารมณ์ที่ถูกต้อง (จากการโหวตอาสาสมัคร)	ร่าเริง	47	2	1	5	0	1	47 / 56	83.93 %
	เสียใจ	1	46	2	4	3	1	46 / 57	80.71 %
	โกรธ	7	6	39	4	0	1	39 / 57	68.43 %
	กลัว	1	0	0	53	2	1	53 / 57	92.99 %
	รังเกียจ	1	3	4	5	39	4	39 / 56	69.65 %
	ประหลาดใจ	2	0	2	2	0	50	50 / 56	89.29 %

ตารางที่ 4.14 ตารางแสดงการแจกแจงความแม่นยำของแต่ละอารมณ์
บนวิธีการคัดเลือกแบบเดลต้า-ไคสแควร์

		อารมณ์ที่ทำนาย (จากโปรแกรม)						ความแม่นยำ	ความแม่นยำเป็น %
		ร่าเริง / ดีใจ	เสียใจ	โกรธ	กลัว	รังเกียจ	ประหลาดใจ		
อารมณ์ที่ถูกต้อง (จากการโหวตอาสาสมัคร)	ร่าเริง / ดีใจ	52	1	0	2	0	1	52 / 56	92.86 %
	เสียใจ	0	50	1	6	0	0	50 / 57	87.72 %
	โกรธ	0	1	46	8	2	0	46 / 57	80.71 %
	กลัว	0	0	0	57	0	0	57 / 57	100 %
	รังเกียจ	0	4	1	5	45	1	45 / 56	80.36 %
	ประหลาดใจ	0	0	2	2	0	52	52 / 56	92.86 %

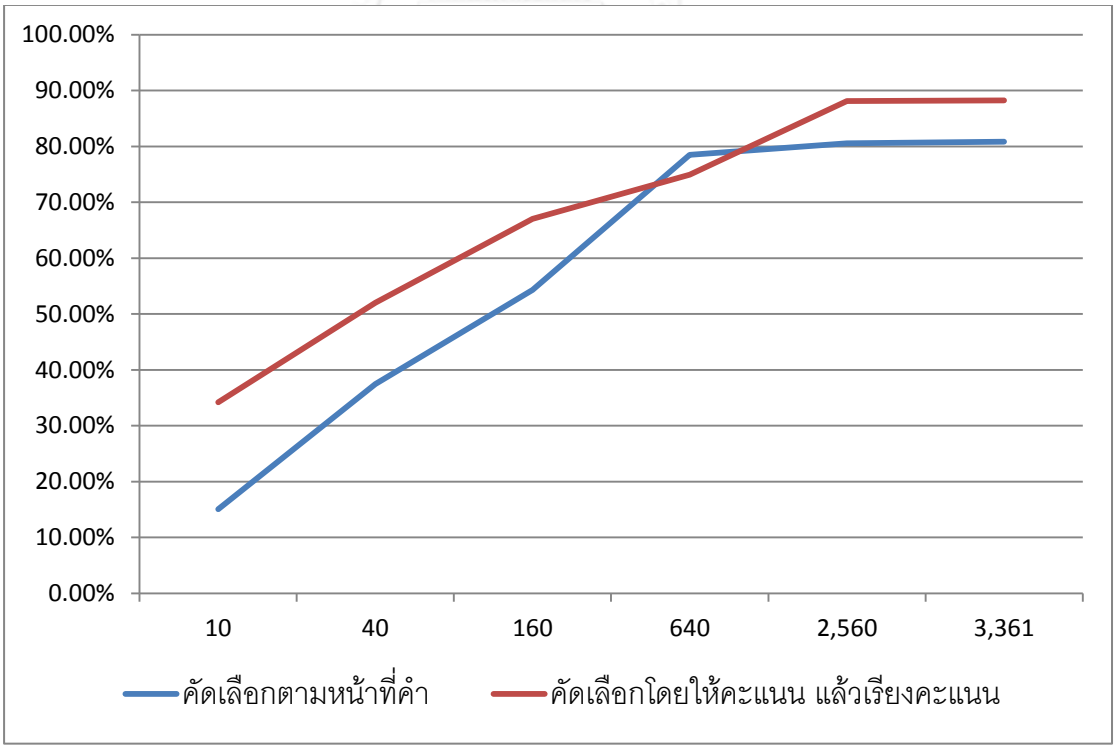
จากตารางที่ 4.13-4.14 ตารางแสดงการแจกแจงความแม่นยำของแต่ละอารมณ์บนวิธีการคัดเลือกระหว่างไคสแควร์ กับ เดลต้า-ไคสแควร์ พบว่าความแม่นยำบนวิธีการคัดเลือกแบบเดลต้า-ไคสแควร์ ให้ผลความแม่นยำที่เพิ่มขึ้นในแต่ละอารมณ์ เทียบกับ บนวิธีการคัดเลือกแบบไคสแควร์ โดยที่ผลความแม่นยำที่เพิ่มขึ้นมากที่สุด คือ 12.28% บนข้อความอารมณ์โกรธ และ ผลความแม่นยำที่เพิ่มขึ้นน้อยที่สุด คือ 3.57% บนข้อความอารมณ์ประหลาดใจ ตามลำดับ

4.3. การทดลองความแม่นยำของวิธีการคัดเลือกของคำ

ในการทดลองนี้ จะทำการหาความแม่นยำของวิธีการคัดเลือกของคำ เปรียบเทียบระหว่างวิธีการคัดเลือกตามหน้าที่คำ (Part of Speech, POS) กับ คัดเลือกโดยให้คะแนนแล้วเรียงคะแนน ในส่วนวิธีการคัดเลือกตามหน้าที่คำที่เคยนำเสนอในงานวิจัยที่ผ่านมา บนข้อความภาษาไทย [2] คัดเลือกหน้าที่ของคำ คือ คำนาม (Noun) กับ คำกริยา (Verb) จากการทดลองที่แล้ว ชุดของคำประกอบด้วย คำนามจำนวน 2,102 คำ และ คำกริยา 1,259 คำ ดังนั้นจำนวนคำนาม กับ คำกริยา มีจำนวนทั้งสิ้น 3,361 คำ ดังแสดงตารางที่ 4.15

ตารางที่ 4.15 การทดสอบความแม่นยำ บนข้อมูลชุดทดสอบ
เปรียบเทียบระหว่างวิธีการคัดเลือกตามหน้าที่คำ กับ คัดเลือกโดยให้คะแนนแล้วเรียงคะแนน

วิธีการคัดเลือก คุณลักษณะของคำ	จำนวนคำ					
	10	40	160	640	2,560	3,361
คัดเลือกตามหน้าที่คำ (Part of Speech, POS)	15.04%	37.46%	54.28%	78.47%	80.53%	80.83%
คัดเลือกโดยให้คะแนนแล้วเรียงคะแนน	34.21%	52.00%	67.00%	74.93%	88.13%	88.20%



ภาพที่ 4.5 การทดสอบความแม่นยำ บนข้อมูลชุดทดสอบ
เปรียบเทียบระหว่างวิธีการคัดเลือกตามหน้าที่คำ กับ คัดเลือกโดยให้คะแนนแล้วเรียงคะแนน

จากตารางที่ 4.15 วิธีการคัดเลือกทั้งสองวิธี ใช้วิธีการให้น้ำหนักคำแบบเดลต้า-ไคสแควร์ แล้วทำการเพิ่มจำนวนคำเป็น 10, 40, 160, 640, 2,560 และ 3,361 คำ เพื่อพิจารณาการเปลี่ยนแปลงของความแม่นยำในการจำแนกข้อความ พบว่าทั้งสองวิธีเมื่อจำนวนคำเพิ่มขึ้นทำให้ความแม่นยำเพิ่มขึ้นตามด้วย แต่เมื่อจำนวนคำที่เพิ่มขึ้นเข้าใกล้จำนวนคำทั้งหมดเท่าไร ทำให้อัตราการเพิ่มความแม่นยำจะลดลงเท่านั้น เมื่อเปรียบเทียบกับจำนวนที่น้อยกว่า และ วิธีคัดเลือกโดยให้คะแนนแล้วเรียงคะแนน จะให้ผลความแม่นยำที่ดีกว่า วิธีการคัดเลือกตามหน้าที่คำ โดยที่ผลความแม่นยำสูงสุดของวิธีคัดเลือกตามหน้าที่คำกับ คัดเลือกโดยให้คะแนนแล้วเรียงคะแนน เท่ากับ 80.83% กับ 88.20% ตามลำดับ วิธีที่นำเสนอประสิทธิภาพความแม่นยำเพิ่มขึ้นอยู่ 7.37%

4.4. การทดลองเปรียบเทียบความแม่นยำบนชุดข้อความระหว่างเฟสบุ๊ค กับ พันทิป

ในการทดลองนี้ จะทำการทดลองเปรียบเทียบความแม่นยำบนชุดข้อความระหว่างเฟสบุ๊ค [7] กับ พันทิป [8] เพื่อพิจารณาแหล่งที่มาของข้อความว่า มีผลกับความแม่นยำในการจำแนกข้อความหรือไม่ คัดเลือกข้อความจากการโหวตอารมณ์ของอาสาสมัคร เหลือจำนวน 3,390 ข้อความ ประกอบด้วย ข้อความมาจากพันทิป จำนวน 2,502 ข้อความ และข้อความมาจากเฟสบุ๊ค จำนวน 888 ข้อความนำมาสร้างโมเดลสำหรับฝึกฝน โดยใช้วิธีการหาฟังก์ชันความหนาแน่นของความน่าจะเป็น สำหรับการกระจายตัวเกาส์เซียนแบบปกติ (Probability Density Function for Normal Gaussian Distribution) [10] และใช้อัลกอริทึมของนาอีฟเบย์ (Naïve Bayes algorithm) [10]

ตารางที่ 4.16 การทดสอบความแม่นยำ บนข้อมูลชุดทดสอบ
เปรียบเทียบแหล่งชุดข้อความระหว่าง เฟสบุ๊ค กับ พันทิป

แหล่งชุดข้อความ	การให้น้ำหนัก	
	$tf.\chi 2$	$tf.\delta\text{-}\chi 2$
เฟสบุ๊ค	80.68%	84.09%
พันทิป	79.60%	89.77%

จากตารางที่ 4.16 แสดงผลความแม่นยำ เปรียบเทียบแหล่งชุดข้อความระหว่างเฟสบุ๊ค กับ พันทิป พบว่าความแตกต่างความแม่นยำของแหล่งชุดข้อความระหว่าง เฟสบุ๊ค กับ พันทิป เท่ากับ 1.08% บนวิธีการให้น้ำหนักแบบ $tf.\chi 2$ และ เท่ากับ 5.68% บนวิธีการให้น้ำหนักแบบ $tf.\delta\text{-}\chi 2$ ในส่วนวิธีการให้น้ำหนักแบบ $tf.\delta\text{-}\chi 2$ จะให้ผลความแม่นยำที่ดีกว่า $tf.\chi 2$ ทั้งแหล่งชุดข้อความเฟสบุ๊ค กับ พันทิป โดยผลความแม่นยำที่เพิ่มขึ้นเท่ากับ 3.41% บนเฟสบุ๊ค กับ 10.17% บนพันทิป ตามลำดับ

บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

5. สรุปผลการวิจัย และข้อเสนอแนะ

5.1. สรุปผลการวิจัย

งานวิจัยนี้ นำเสนอวิธีการคัดเลือก และ การให้น้ำหนักของคำที่มีคุณลักษณะเชิงอารมณ์ บนชุดข้อความจากเครือข่ายสังคมออนไลน์ ได้แก่ เฟสบุ๊ค กับ พันทิป จากการทดลองสามารถสรุปเป็นตารางได้ดังตารางที่ 5.1 ที่แสดงด้านล่าง

ตารางที่ 5.1 ตารางผลสรุปวิธีการระหว่าง งานวิจัยที่ผ่านมา กับ งานวิจัยนี้

	วิธีการที่ใช้ในงานวิจัยที่ผ่านมา	**วิธีการที่นำเสนอในงานวิจัยนี้**	**ผลสรุป**		**ตารางอ้างอิง**
			ความแม่นยำ	**เหตุผล**	
การให้น้ำหนักของคำ	การให้น้ำหนักของคำ แบบความถี่ของคำ กับ ไคสแควร์ ($tf.\chi 2$)	การให้น้ำหนักของคำ แบบความถี่ของคำ กับ เดลต้า-ไคสแควร์ ($tf.\delta$-$\chi 2$)	เพิ่มขึ้น 7.96%	การให้น้ำหนักแบบ $tf.\chi 2$ จะให้น้ำหนักทุกคำบนข้อความ ขณะที่แบบ $tf.\delta$-$\chi 2$ จะให้น้ำหนักบางคำบนข้อความ ผ่านพารามิเตอร์เดลต้า (δ)	ตารางที่ 4.2, 4.7
การคัดเลือกคำ	การคัดเลือกตามหน้าที่ของคำ (Part of Speech, POS) คือ คำนาม กับ คำกริยา	การคัดเลือกตามอันดับคะแนนของคำ	เพิ่มขึ้น 7.37%	การคัดเลือกตามหน้าที่ของคำ จะคัดเลือกเฉพาะคำนาม กับคำกริยา ขณะที่การคัดเลือกตามการให้คะแนน จะคัดเลือก บนทุกหน้าที่ของคำ ตามที่วิเคราะห์ตาราง 4.3-4.4	ตารางที่ 4.3-4.6
การเรียนรู้แบบมีผู้สอน	การเรียนรู้แบบมีผู้สอน โดยการโหวตทั้ง เจ้าของ และไม่ใช่เจ้าของข้อความ	การเรียนรู้แบบมีผู้สอน โดยการจับคู่ระหว่างเจ้าของ กับไม่ใช่เจ้าของข้อความ	ลดลง 2.36%	เนื่องจากการเรียนรู้แบบมีผู้สอน โดยการจับคู่ จะเป็นการอิงเพียงเจ้าของข้อความ แต่ขณะที่การเรียนรู้แบบมีผู้สอน โดยการโหวต จะเป็นการอิงทั้งเจ้าของข้อความ และไม่ใช่เจ้าของข้อความอย่างเท่าเทียมกัน	ตารางที่ 4.8

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 5.1 สามารถสรุปหัวข้อได้เป็นสามส่วนหลัก คือ การให้น้ำหนักของคำ, การคัดเลือกคำ และ การเรียนรู้แบบมีผู้สอน โดยมีการเปรียบเทียบระหว่างวิธีการที่ใช้ในงานวิจัยที่ผ่านมากับ วิธีการที่นำเสนอในงานวิจัยนี้ ในส่วนการให้น้ำหนักของคำแบบเดลต้า-ไคสแควร์ จะให้ผลความแม่นยำที่เพิ่มขึ้นเมื่อเทียบกับการให้น้ำหนักของคำแบบไคสแควร์ อยู่ที่ 7.96% ในส่วนการคัดเลือกตามอันดับคะแนนของคำ จะให้ผลความแม่นยำที่เพิ่มขึ้นเมื่อ เทียบกับการคัดเลือกตามหน้าที่ของคำ อยู่ที่ 16.45% และ ในส่วนการเรียนรู้แบบมีผู้สอน โดยการจับคู่ เทียบกับ การโหวตจากอาสาสมัครให้ผลความแม่นยำลดลง อยู่ที่ 2.36%

จากการทดลองจึงสรุปได้ว่า ทั้งวิธีการคัดเลือกตามหน้าที่คำ และ คัดเลือกโดยการให้คะแนนแล้วเรียงคะแนน เมื่อจำนวนคำเพิ่มขึ้นทำให้ความแม่นยำในการจำแนกเพิ่มขึ้นตามด้วย แต่เมื่อจำนวนคำที่เพิ่มขึ้นเข้าใกล้จำนวนคำทั้งหมดเท่าไร จะทำให้อัตราการเพิ่มความแม่นยำจะลดลงเท่านั้น ส่วนความแม่นยำในการจำแนกของคำที่มีการเรียงคะแนนจำนวน 2,560 กับ 4,260 คำ จะพบว่ามีค่าความแม่นยำที่ใกล้เคียงกัน นั้นแสดงให้เห็นว่า จำนวน 2,560 คำที่มีการเรียงคะแนนของคำ มีประสิทธิภาพการจำแนกอารมณ์บนข้อความ เทียบเท่ากับจำนวนคำทั้งหมด 4,260 คำ สุดท้ายคือวิธีการคัดเลือก และ การให้น้ำหนักของคำแบบเดลต้า-ไคสแควร์ (δ-$\chi 2$) จะให้ผลความแม่นยำที่ดีขึ้นเมื่อเทียบกับวิธีอื่น เนื่องจากวิธีดังกล่าวจะพยายามคัดคำที่มีคุณลักษณะที่เกิดขึ้นในหลายอารมณ์ออก ให้คงเหลือชุดของคำที่มีคุณลักษณะที่เกิดขึ้นเฉพาะในอารมณ์มากขึ้น

5.2. ข้อเสนอแนะ

เนื่องจากงานวิจัยนี้ นำเสนอวิธีการคัดเลือก และ การให้น้ำหนักของคำแบบเดลต้า-ไคสแควร์ (δ-$\chi 2$) ซึ่งเป็นวิธีการเชิงสถิติเพื่อคัดเลือกคำที่มีคุณลักษณะที่เกิดขึ้นในอารมณ์เฉพาะเข้ามา ดังนั้นจึงต้องใช้ปริมาณข้อความจำนวนมาก เพื่อยืนยันให้ได้ว่าคำนั้นมีคุณลักษณะที่เกิดขึ้นในหลายอารมณ์

ในส่วนความคลอบคลุมของคำที่เกิดขึ้น เนื่องจากงานวิจัยนี้ คัดเลือกข้อความที่เจ้าของข้อความสามารถมาระบุอารมณ์ได้เท่านั้น ดังนั้นจึงมีความเป็นไปได้ว่าคำที่เกิดขึ้นบนชุดข้อความจะไม่คลอบคลุมกับคำที่เกิดขึ้นจริงบนเครือข่ายสังคมออนไลน์

ปัญหาที่คาดว่าจะเกิดขึ้นตามมาจากการคัดเลือกคำ คือ แม้ว่าจะรับข้อความมาเป็นจำนวนมาก แต่อาจจะยืนยันไม่ได้ว่าคำนั้นมีคุณลักษณะที่เกิดขึ้นในหลายอารมณ์หรือไม่ เพราะมีความเป็นไปได้ว่าคำนั้นอาจจะเกิดขึ้นน้อยครั้ง แต่เป็นคำที่มีคุณลักษณะที่เกิดขึ้นในอารมณ์เฉพาะ อีกทั้งยังมีคำที่ยังไม่เคยเกิดปรากฏบนข้อมูลชุดฝึกฝน ซึ่งมีความเป็นไปได้ เพราะในเครือข่ายสังคมออนไลน์ มีคำที่เกิดขึ้นใหม่อยู่บ่อยครั้ง ดังนั้นงานวิจัยในอนาคต คือ การคัดเลือกคำที่มีคุณลักษณะเชิงอารมณ์ และสามารถยืนยันได้ว่าคำนั้นเป็นคุณลักษณะเฉพาะของอารมณ์นั้นจริง แม้ว่าคำนั้นจะเกิดขึ้นน้อย และยังไม่เกิดขึ้นจริง

มีแนวคิดเรื่องการสร้างข้อความหรือประโยคจากชุดของคำที่มีอยู่ ซึ่งถ้าได้ชุดของคำที่คลอบคลุมแล้ว จะสามารถสร้างข้อความหรือประโยคที่คลอบคลุมชุดของคำนั้นได้ อีกทั้งยังสามารถควบคุมการเกิดขึ้นของคำต่างๆ ได้อีกด้วย แล้วนำวิธีการคัดเลือกคำจากงานวิจัยนี้ มาประยุกต์ใช้กับชุดข้อความที่สร้างขึ้นมา เพื่อให้ได้ชุดของคำที่มีคุณลักษณะเฉพาะของอารมณ์นั้น และ นำไปจำแนกอารมณ์บนข้อความเพื่อเตรียมชุดข้อความสำหรับเทคโนโลยีการรู้จำอารมณ์ของมนุษย์ (Human Emotion Recognition, HER) ต่อไป

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รายการอ้างอิง

1. Danisman, T., *Feeler: Emotion Classification of Text Using Vector Space Model.* In AISB 2008 Convention, Communication, Interaction and Social Intelligence, Affective Language in Human and Machine, vol.2, 2008.
2. Inrak, P., *Applying Latent Semantic Analysis To Classify Emotions In Thai Text.* 2nd International Conference on Computer Engineering and Technology, 2010.
3. Lee, C., *Emotion Recognition for Affective User Interfaces using Natural Language Dialogs.* 16th IEEE International Conference on Robot & Human Interactive Communication, Jeju, Korea, August 26-29, 2007.
4. DAI, L., *Using Modified CHI Square and Rough Set for Text Categorization with Many Redundant Features.* 2008 International Symposium on Computational Intelligence and Design, 2008.
5. Lan, M., *Supervised and Traditional Term Weighting Methods for Automatic Text Categorization.* IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2009.
6. Debole, F., *Supervised Term Weighting for Automated Text Categorization.* Proc. ACM Symp. Applied Computing, 2003: p. 784-788.
7. Facebook, [Online] http://www.facebook.com/, [Accessed 8 October 2012].
8. Pantip, [Online] http://www.pantip.com/, [Accessed 8 October 2012].
9. Ekman, P., *Facial Expressions of Emotion.* American Psychological Association, Apl. 1993, 1992: p. 376-379.
10. Buntine, W.L., *Operations for Learning with Graphical Models.* Journal of Artificial Intelligence Research 2, Dec. 1994, 1994: p. 159-225.
11. A., T., *Impact Parameter on LSA Performance for Thai Text Summarization.* Kaset Vichakarn'43, Bangkok, Thailand, 2004.
12. Haruechaiyasak, C., *TLex: Thai Lexeme Analyser Based on the Conditional Random Fields.* 2009.
13. Yang, Y., *A Comparative Study on Feature Selection in Text Categorization.* Proc. Int'l Conf. Machine Learning, 1997: p. 412-420.
14. Docs, G., [Online] http://www.docs.google.com/, [Accessed 10 October 2012].
15. KK•Spell, [Online] http://www.kkspell.com/, [Accessed 6 November 2012].
16. Microsoft Corporation, M.V.S., [Online] http://www.microsoft.com/visualstudio/th-th/download, [Accessed 8 November 2012].

17. 1999-2012, W.V.c., [Online] http://www.cs.waikato.ac.nz/~ml/weka/, [Accessed 8 November 2012]. **The University of Waikato, Hamilton, New Zealand**.
18. Sornlertlamvanich, V., *ORCHID: Thai Part-Of-Speech Tagged Corpus.* Technical Report Orchid TRNECTEC-1997-001, National Electronics and Computer Technology Center, Thailand, 1997.
19. L., D.R., *Emotional expression and brain physiology. I: Approach/withdrawal and cerebral asymmetry.* Journal of Personality and Social Psychology, 58, 1990: p. 330-341.
20. D., B.J., *Recognition of facial expression in three cultures. Journal of Cross-cultural Psychology.* 1980: p. 263-280.
21. W., R., *Will the real relationship between facial expression and affective experience please stand up: the case of exhilaration.* Cognition and Emotion, 1995: p. 33-58.
22. Kruengkrai, C., *A Conditional Random Field Framework for Thai Morphological Analysis.* Proc. of the Fifth Int. Conf. on Language Resources and Evaluation (LREC-2006), 2006.
23. Khoo, C.S.G., *Using Statistical and Contextual Information to Identify Two-and Three-Character Words in Chinese Text.* Journal of the American Society for Information Science and Technology, 2002: p. 365–377.
24. Gil, J., *Projections in Venn-Euler Diagrams.* Visual Languages, 2000. Proceedings. 2000 IEEE International Symposium, 2000: p. 119-126.
25. Salton, G., *A vector space model for automatic indexing, Communications of the ACM.* 1975: p. 613-620.
26. Huang, J., *The Comparison of Classifiers for Object Categorization Based on Bag-of-Word Technology.* Pattern Recognition, 2009. CCPR 2009. Chinese Conference on 4-6 Nov. 2009, 2009.
27. Salton, G., *Term-Weighting Approaches in Automatic Text Retrieval.* Information Processing and Management, vol. 24, no. 5, 1988: p. 513-523.
28. Silverman, B.W., *Density Estimation for Statistics and Data Analysis.* Published in Monographs on Statistics and Applied Probability, London: Chapman and Hall, 1986.
29. John, G., *Estimating Continuous Distributions in Bayesian Classifiers.* In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, 1995.
30. Thomas, B., *An Essay towards solving a Problem in the Doctrine of Chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, A.*

M. F. R. S. Philosophical Transactions of the Royal Society of London 53, 1763: p. 370–418.

31. Dawid, A.P., *Conditional Independence in Statistical Theory.* Journal of the Royal Statistical Society, 1979: p. 1–31.
32. Deng, Z.-H., *A Comparative Study on Feature Weight in Text Categorization.* Proc. Asia-Pacific Web Conf., vol. 3007, 2004: p. 588-597.
33. Charoenpornsawat, P., *Feature-based Thai Word Segmentation. Master's Thesis. Computer Engineering.* Chulalongkorn University, Bangkok, Thailand, 1999.
34. T.Mitchell, *Machine Learning.* McGraw Hill, 1997.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ก
รหัสหน้าที่ของคำใน ORCHID

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

No.	POS	Description	Example
1	NPRP	Proper noun	วินโดวส์ 95, โคโรน่า, โค้ก, พระอาทิตย์
2	NCNM	Cardinal number	หนึ่ง, สอง, สาม, 1, 2, 3
3	NONM	Ordinal number	ที่หนึ่ง, ที่สอง, ที่สาม, ที่1, ที่2, ที่3,
4	NLBL	Label noun	1, 2, 3, 4, ก, ข, a, b
5	NCMN	Common noun	หนังสือ, อาหาร, อาคาร, คน
6	NTTL	Title noun	ดร., พลเอก
7	PPRS	Personal pronoun	คุณ, เขา, ฉัน
8	PDMN	Demonstrative pronoun	นี่, นั่น, ที่นั่น, ที่นี่
9	PNTR	Interrogative pronoun	ใคร, อะไร, อย่างไร
10	PREL	Relative pronoun	ที่, ซึ่ง, อัน, ผู้
11	VACT	Active verb	ทำงาน, ร้องเพลง, กิน
12	VSTA	Stative verb	เห็น, รู้, คือ
13	VATT	Attributive verb	อ้วน, ดี, สวย
14	XVBM	Pre-verb auxiliary, before negator “ไม่”	เกิด, เกือบ, กำลัง
15	XVAM	Pre-verb auxiliary, after negator “ไม่”	ค่อย, น่า, ได้
16	XVMM	Pre-verb, before or after negator “ไม่”	ควร, เคย, ต้อง
17	XVBB	Pre-verb auxiliary, in imperative mood	กรุณา, จง, เชิญ, อย่า, ห้าม
18	XVAE	Post-verb auxiliary	ไป, มา, ขึ้น

No.	POS	Description	Example
19	DDAN	Definite determiner, after noun without classifier in .	นี่, นั่น, โน่น, ทั้งหมด
20	DDAC	Definite determiner, allowing classifier in .	นี้, นั้น, โน้น, นู้น
21	DDBQ	Definite determiner, between noun and classifier or preceding quantitative expression	ทั้ง, อีก, เพียง
22	DDAQ	Definite determiner, following quantitative	พอดี, ถ้วน
23	DIAC	Indefinite determiner, following noun; allowing classifier in .	ไหน, อื่น, ต่างๆ
24	DIBQ	Indefinite determiner, between noun and classifier or preceding quantitative expression	บาง, ประมาณ, เกือบ
25	DIAQ	Indefinite determiner, following quantitative	กว่า, เศษ
26	DCNM	Determiner, cardinal number expression	หนึ่งคน, เสือ 2 ตัว
27	DONM	Determiner, ordinal number expression	ที่หนึ่ง, ที่สอง, ที่สุดท้าย
28	ADVN	Adverb with normal form	เก่ง, เร็ว, ช้า, สม่ำเสมอ
29	ADVI	Adverb with iterative form	เร็วๆ, เสมอๆ, ช้าๆ
30	ADVP	Adverb with prefixed form	โดยเร็ว
31	ADVS	Sentential adverb	โดยปกติ, ธรรมดา
32	CNIT	Unit classifier	ตัว, คน, เล่ม
33	CLTV	Collective classifier	คู่, กลุ่ม, ฝูง, เชิง, ทาง, ด้าน, แบบ, รุ่น
34	CMTR	Measurement classifier	กิโลกรัม, แก้ว, ชั่วโมง

No.	POS	Description	Example
35	CFQC	Frequency classifier	ครั้ง, เที่ยว
36	CVBL	Verbal classifier	ม้วน, มัด
37	JCRG	Coordinating conjunction	และ, หรือ, แต่
38	JCMP	Comparative conjunction	กว่า, เหมือนกับ, เท่ากับ
39	JSBR	Subordinating conjunction	เพราะ, ว่า, เนื่องจาก, ที่, แม้ว่า, ถ้า
40	RPRE	Preposition	จาก, ละ, ของ, ใต้, บน
41	INT	Interjection	โอ้ย, โอ้, เออ, เอ๋, อ๋อ
42	FIXN	Nominal prefix	การทำงาน, ความสนุกสนาน
43	FIXV	Adverbial prefix	อย่างเร็ว
44	EAFF	Ending for affirmative sentence	จ๊ะ, จ้ะ, ค่ะ, ครับ, นะ, น่า, เถอะ
45	EITT	Ending for interrogative sentence	หรือ, เหรอ, ไหม, มั๊ย
46	NEG	Negator	ไม่, มิ, ได้, ไม่, ได้, มิ
47	PUNC	Punctuation	(,),", ,, ;

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ข
ตัวอย่างข้อความภาษาไทย จำนวน 300 ข้อความ
จากเครือข่ายสังคมออนไลน์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ข้อความที่	ตัวอย่างข้อความ
1	วีระ เพื่อน BCC148 นายเป็นเพื่อนที่ดีมากคนนึงเลยที่เรารู้จัก วันนี้ไม่ว่านายอยู่ที่ไหน เราเป็นเพื่อนกันเสมอนะ RIP หลับให้สบายนะเพื่อน
2	จริงๆถ้าจะตั้งสปอยแต่ละตอนก็น่าจะไปตั้งในกระทู้เรื่องนี้ที่ปักหมุดไว้ดีกว่านะ ดีกว่ามาตั้งห้องสปอยทีละตอนทุกอาทิตย์ มันรก!!!
3	ตอนล่าสุด นั่งดูอยู่ขนลุกเลย เพราะ ตอนที่กล้องล้มอ่ะค่ะ ว่าตั้งนานไม่ล้ม อ่ะ เห็นเงาขาวแว๊บไปเหมือนผู้หญิง อ่ะ
4	หากไม่คืนความตายแก่ผู้ตายคำสาปก็ไม่หยุดลง หากคำสาปยังคงดำเนินไปจนจบการศึกษาแล้ว ผู้ตายในปีนั้นจะตายหรือมีชีวิตอยู่ต่อไปเหรอ
5	ลูกเสือผมอาบวันแรก น้ำในบ่อมันยังสะอาด พอวันที่สอง มันเล่นโดดลงบ่อที่เค้าให้ตักอาบจนมันดำ แต่ก็ต้องทนอาบ สามวันอาบน้ำน้อยมาก
6	ผมไม่เคยเชื่อว่าระบบยุติธรรมไทยจะสามารถส่งนักการเมืองที่สร้างเครือข่ายปกป้องตัวเอง ให้เอาเข้าคุกได้ อย่างมากก็รอลงอาญา
7	ใช่ เธอเป็นคนบอกรักฉันทำให้ฉันดีใจและมีความสุขมากและก็เป็นคนเดียวกันที่บอกเลิกฉันและทำให้ฉันมีน้ำตาและเสียใจที่สุด
8	พอเปิดให้เข้างานเท่านั้นแหละ แทนที่จะเดินกันเป็นแถวให้มันเรียบร้อยน่ารัก = = เจือกวิ่งกันหน้าตั้ง จะให้ยืนเฉยก็เกรงจะชวด
9	เห็นเรื่องคนโดนไฟครอกแบบนี้แล้ว นึกถึงเรื่อง silent hill เด็กผู้หญิงโดนย่างในเรื่อง แต่เรื่องนี้ดูน่าสยองกว่า
10	จริงลองเซิร์ชดูแล้วที่เค้าปล่อยโหลดกันชื่อว่า Bakemonogatari 5.5 Special Recap ครับ มีใครพอทราบมั้ยครับว่ามันคืออะไร?
11	ที่นั่งเบาะแคบและแข็งไปนิดตามสไตล์อเมริกัน ที่รองศีรษะยื่นออกมาทำให้เมื่อยคอมากๆๆๆ ไม่เหมาะกับรูท long haul สำหรับเรา -_-"
12	อยากไปล่าท้าผี..มากๆ...อยากรู้..อยากลอง เปนคนมีสัมพัสพิเศษเหมือนกัน..อยากรู้ว่าเราคิดไปเองหรืออะไร..ยังไง
13	ทำใจกันได้แล้วเรื่องนี้ เพราะผลโพลล์ชั้นนำออกมาว่า เด็กสมัยนี้มากกว่า 50%คิดว่าเรื่องนี้เป็นเรื่องธรรมดารับกันได้ครับ
14	มันตั้ง4ปีคับทำไมยังไม่เสร็จมีอะไรก็น่าจะออกมาประกาศกันตั้งแต่ช่วงแรกๆมาประกาศตอนนี้เหมือนต่างคนต่างปัดความรับผิดชอบ
15	ตอนจบอ่ะ..ที่เจมส์กับเชอร์รี่เดินออกมา..สังเกตไม๊มีอะไรอยู่ในห้องข้างหลัง..ดูดีๆ..เอ๊ะหรือว่าเราตาฝาดหว่า..
16	แผ่น 1 ที่โน่นข้าพเจ้ารอดตัวไป ไม่มี PR-WS SAO ถ้ามีแถมมานี้ถึงกับลงไปกองกับพื้นแบบไม่รู้สาเหตุแน่ๆ ฮ่า~

17	ช่างแม่งเหอะ แม่งก็เหมือนเดิม ถามหน่อยเหอะในภาคีมีใครเลิกดูช่อง3รึยัง ปากก้อว่าอย่างแต่กลับบ้าน"นอนดูแรงเงา" ถุย!!!!!!
18	ถ้า president ไม่ยอมกลับไปใช้กฎบัตรเก่าเด็กรุ่นต่อๆไปรวมถึงที่เพิ่งจบก็จะไม่สามารถเอาวุฒิไปใช้สมัครงานได้
19	ผมว่าไม่ไม่ใช่ text สปอยเลยอะปกติพวกtextแบบนี้มันจะมีคำบรรยายรายละเอียดมากกว่านี้ด้วยนี้อะไรไม่รู อ่านไปมีนิดเดียวเอง
20	มันกำลังจะให้ประเทศไทยใช้เงินของหลวงตาล่ะมั้ง ยังเหลือเยอะนี่ ใครจะช่วยป้องกันได้มั่งอ่ะ ควันออกหูเหมือนกันแระเรา
21	ดูแล้วค่ะ สยองฉากที่ เค้าเรียกว่าไงอ่ะ ประมาณลอกชั้นผิวหนังที่เสียออกในโรงบาลอ่ะคะ ทำได้เหมือนอ่ะ น่ากลัวว
22	ท่าน ค่ะ เมื่อไหร่เวรกรรม มันจะตามมาทัน วัฒนา อัศวเหม เสนาะ เทียนทอง บรรหาร ศิลปอาชา ฯลฯ ในรัฐบาลชุดนี้
23	เราเริ่มแล้ว สินค้าใครมันยังสนันสนุนมัน เนาก็ไม่ซื้อมันและเพื่อนๆเราก็เริ่มแล้ว ดูซิมันยังด้านสนันสนุนคนด้านอยู่อีกไหม
24	เจ้าเหมียวน้อย แกเป็นแมวที่แข็งแกร่งมากรู้มั้ย ไม่ว่าแกจะล้มกี่ครั้งก็ยังพยายามลุกเพื่อต่อสู้กับโลกใบนี้
25	คนที่คุยกันแค่ครั้งเดียวกลับช่วยเหลือเราเต็มที่ ต่างกับ O.P ที่คุยกันหลายครั้งแต่กลับไม่ช่วยอะไรเราเลย เฮ้อเหนื่อยใจ
26	ทีประเทศไทยเป็นแบบนี ก็เพราะนักากรเมืองไทยรัฐบาลไทยต้องการให้คนไทยโง่ๆ เข้าไว้ โง่เอาไว้ จะได้ปกครองง่ายยังไงหละ
27	ฉากเครื่องตัดหญ้า ที่เราเห็นรางๆแค่เห็นว่าเครื่องตัดหญ้าไถไปตามพื้นหญ้าแล้วก็มีเสียงดังเราสะดุ้ง
28	หงายเงิบ ตรงคำว่า นกกะ(จอก) เนี่ยละ จอกศักดิ์สิทธิ์ จริงๆ -*- แต่ว่าภาคค้างคืน อยู่คนละค่ายนิ -.-
29	เป็นผู้สื่อข่าวต้องมีจริยธรรมสูง หันไปเป็นนักการเมืองพวกเดียวกันดีกว่า จะได้โกงกันมากกว่านี้ บ้านเมืองจะฉิบหายอยู่แล้ว
30	มันทำให้รู้เลยว่าขนาดแมวตัวเล็กๆมันยังสู้เพื่อจะมีชีวิตอยู่เลย เมื่อมีปัญหาฉันจะนึกถึงแกนะเจ้าเหมียวน้อย
31	2 วันนี้ผมเจอเรื่องเศร้าถึง 2 เรื่อง วันนี้ขอใช้ wall ของผมเพื่อไว้อาลัยให้กับเพื่อนและเพื่อนร่วมโลกของผมด้วยนะครับ
32	ผมน้อยใจมากเลยเพราะปากคนๆนี้ ถ้าเป็นสื่อแล้วรู้ไม่จริง นักข่าวจอมปากเสียสร้างภาพ เดี๋ยวคนก็จะเข้าใจความจริงเอง
33	พวกฝูงควาย มันกำลังเข้ามาหากินหญ้าอีกแล้ว หญ้าในเมืองสีม่วง สีเทาไง แม่ง ธรณีสูบมันลงไปนรกซักที
34	อ่านจบแล้ว อึ้งนิดๆ แต่ก็รอดูตอนต่อไปว่าทำไมอากิระแกถึงได้หลุดรอดมาได้และทำไปเพื่ออะไร = ="

35	อาจจะใช่ก็ได้ คนไทยส่วนใหญ่ไม่ค่อยเคารพกฎสังคมสักเท่าไร เหมือนกฎหมาย พอทำผิด ก็ไม่เอาเรื่อง ปัดความรับผิด แถ ต่างๆนาๆ
36	เรื่องนี้สุดยอดมาก ทั้งภาพ ทั้งเสียง เนื้อเรื่อง ตัวละคร ทุกอย่างเป็นเหตุเป็นผลกันหมดเลย ลงตัวทุกอย่าง
37	ผลกระทบจะเกิดกับใครก็ไม่คำนึงถึง มันเริ่มรุนแรงขึ้นเรื่อยๆ จะชอบจะรักอะไรก็ต้องไตร่ตรอง และก็ไม่รู้ว่าจะถูก หรือผิด
38	แค่เรื่องความเสียหายของจำนำข้าวกับน้ำท่วม ก็น่าอับอายจนควรลาออกจากการเป็นส.สทั้งพรรค และทั้งคณะรัฐมนตรีแล้ว!
39	แอบรู้สึกถึงลางร้าย อาสึกะตรูจะรอดมั้ยเนี่ย ยิ่งเฟลๆแทนชินจิ Forever Alone ในมังงะอยู่ TT^TT
40	ผมว่าเหมือนโอบิโตะจะอัพเกรดเก่งเวอร์ไปไหมหว่าตอนนี้ แบบพึ่งทำกายภาพบำบัดฟื้นตัวมาได้เอง -_-
41	ละหมาดวันศุกร์เค้าละหมาดที่มัสยิดกันโว้ย ไม่ใช่ที่ไหนก็ได้ ไอปรวก..เหลิม..แค่นี้ไม่รู้มันจบกฎหมายอิสลามมายังไงว่ะ ...
42	ทำให้เด็กที่เพิ่งจบไปเดือน5-6 ปีนี้และที่กำลังจะจบรุ่นต่อๆไปจะได้รับใบปริญญาบัตรซึ่งไม่ได้รับการรับรอง
43	มายาทั้งนั้น สร้างเรื่องสร้างภาพ ปั้นแต่งจนไม่รู้ว่าตัวจริงของตัวเองเป็นใคร แล้วจะรู้จักความจริงตรงหน้าได้ยังไง
44	อย่าพยามไปเข้าใจพวกมันครับ อย่าทำให้ตัวเองต้องเสื่อมขนาดนั้น ปล่อยพวกมัน ได้แต่มองพวกมันอย่างสังเวชก็พอแล้วครับ
45	เอ๋ ! หลักตอนนี้มันก็แค่คาดเดานี่ คือเหมือนจะมีอะไรหลาย ๆ อย่างที่โยงไปหาโตะกับอาซึนะก็จริง
46	จะว่าไป เซเบอร์ FZ ดูเป็นผู้ใหญ่กว่า FSN อีกนะ -*- ทั้งๆที่เป็นรุ่นก่อนหน้าตั้ง 10ปี -.-
47	น่าอายวะ ผิดก็ยอมรับผิดครับ ลูกผู้ชาย ไม่ชอบเสื้อแดงก็จริง แต่งานนี้ต้องยอมรับผิดโดนโจมตี ก็ต้องยอมรับ อย่าไปเถียง
48	ดูจากภาพรุกป่าจนเตียน ถ้าโดนสึนามิถล่ม หรือน้ำป่าหลาก ก็ไม่ต้องเห็นใจ ช่วยกันสมน้ำหน้าคนโลภไร้ยางอาย
49	นคุณ ragna ย่อมรู้อยู่แก่ใจว่าตัวคุณ คือ กระต่าย หรือ หมาป่า ฉะนั้น อย่ามาลงอารมแต่เราเพราะเรื่องประมูลเป็นหลัก
50	ทำแบบนี้ประชาชนไม่มีที่ทำกินก็ต้องไปพึงพิงเขา เป็นลูกจ้างเขา มันก็จะเป็นวัฏจักรที่คนรวยก็รวยเอาๆ คนจนก็จนต่อไป
51	โคดดดดดดด..โหดต่างหาก...หุหุ.ไม่ช่ายแค่นี้นะ...ยังต้องจ่ายเพิ่มอีก..อะไหล่บางตัวยังไม่ได้
52	ประมาณคนประเภทเดียวกันสามารถดึงดูดเข้าหากันได้ คงเป็นอาการแบบแอบๆล่ะนะ รอติดตามกันต่อไป

53	แมร่งมีแต่ปัญหา พวกเหี้ยเนี่ย!! ออกมาเรียกร้องสันตีนอะไรเพื่อเจ้านายที่โกงชาติชิบหายและเพื่อลบล้างความผิดที่เคยก่อ!!
54	ฮันเบจังงง ยอมเสี่ยงคุกเพื่อเธอ!!!!!!
55	คนไทยเจ็ดสืบล้านคน ช่วยเป็นสมาชีกต่อต้านการคอร์รัปชั่น การโกงทุกรูปแบบ การโกง ห้ามถอยเด็ดขาดทำทุกๆวัน
56	จัดหลักสูตรการเรียนการสอนในโรงเรียนรัฐให้มันเสมอภาคกันก่อนเถอะ เด็กในเมืองนอกเมือง การเรียนการสอนห่างกันลิบลับ
57	ยิ่งตอนที่มีภาพผีแว้บๆ โผล่แทรกในกล้องนี่ขนลุกซู่เลย เพราะมันสยองได้โดยไม่ต้องบิ๊วจริงๆ
58	นางเอกเรื่องอื่นทำหนักกว่าไม่มีคนบ่นถึง ฮ่า~ สงสัยเพราะมาแรงอยากให้เค้าสะดุดเล่นกระมัง
59	มันครองเมือ่งไปแระรวมกลุ่มกันไปเดินเกะกะให้ชาวบ้านเขาเดือดร้อนก็ไม่มีตำรวจหน้าไหนกล้าแม่งอยากยิงกะบาลแกนนำมันจิง ๆ
60	ฮาา กะไว้ละต้องมีคนคิดเรื่องนี้เหมือนกัน ตัวเองก็ยังคิดอยู่เลยตอนดู แวบขึ้นมาในสมองเฉยเลย ฮาา ^^
61	ตอนเมียหลวงคัมแบ็คสเตจนี่อยากจะโศกาอาลัยเพราะเป็นกู้ดบายสเตจที่ไปไวมาไวมากค่ะสำหรับสึกุ ฮืออออออ T_T
62	นักการเมืองโกงคอร์รัปชัน ประหารชีวีต ตืดคุกตลอดชีวีต ไม่มีอภัยโทษเด็ดขาด ทีอเมรีกาติดคุกตลอดชีวีต ไม่มีประกัน
63	สยองค่ะ ฉากที่ตัดเอ็นข้อเท้า เราแอบเอื้อมมือไปจับข้อเท้าตัวเองเลย อี๊ออ เสียว >_<
64	แบบAnother ver. นี่ตอนแรกคิดว่าเป็นภาพที่เปลี่ยน ที่แท้เพลงประกอบนี่เอง :lol: :lol:
65	มีลูกบอกลูก มีหลานบอกหลาน จะส่งเข้าเรียนที่ไหนต้องดูให้ดี ดูให้ถึง president เลยว่ามีลับลมคมในเรื่องผลประโยชน์รึไม่
66	กู เกลียด ทักษิณ เกลียดเสื้อแดง เกลียดตำรวจเลวๆ เกลียดพวกหูหนวกตาบอดเกลียดพวกไม่ยอมรับความจริง
67	ภาค TV นี้ ไลบร้า กะ MO3 เลยนะนั่น แอบเสียดายไม่มี Epyon ภาค EW Tallgeese 3 อยู่ฝั่ง กันดั้มไม่ใช่รึ
68	อาจจะไม่ทุกคนนะที่มองอย่างนั้น แต่สำหรับผมที่ทำงานล่ามอยู่ตอนนี้ ก็ยอมรับว่าจริงนะ เรื่องเห็นเป็นเบี้ยล่างน่ะ
69	โรงเรียนผมถ้าอาจารย์เจอโทรศัพท์(ในโรงเรียน)จะริบไปแล้วมาเอาคืนใน 1 ปี ให้หลัง :shock:
70	พระเอกมันไม่ได้เป็นเกย์ครับ แค่เป็นคนดีเกินไป ใจเอ็งทำด้วยอะไรวะ หนักแน่นเป็นบ้า OTL

71	แต่ไอ้เกมส์Rewrite ถ้าจะจีบ คางาริ มันต้องเล่นครบทุกรูทก่อนเลยกลายเป็นต้องจีบตัวคนเลย บอกตามตรงไม่ชอบแบบนี้เลย=0=
72	ยอมรับเลยว่าตอนแรกก็ไม่เก่ง มาเอาดีภาษาอังกฤษได้มากขึ้นก็อีตอนดูซับเมะเป็นอังกฤษนี่ล่ะ :cool:
73	เพราะถึงบอกว่าเราเศร้าเราเสียใจเราทำใจไม่ได้ ก็ไม่ได้อะไรขึ้นมาอยู่ดี มีแต่ทำให้เขาไม่สบายใจเปล่าๆ
74	เลิกสนับสนุน เลิกดูรายการของไอ้บ้านี่มานานแล้ว ข่าวไร้สาระเล่นอยู่ได้เป็นวรรคเป็นเวร ข่าวที่ควรเสนอกลับไม่แตะ
75	ไอ้หลอน ๆ นี่นิดหน่อย จะอยู่ในอารมย์สงสัยซะมากกว่าว่าทำไม เด็กคนนี้ ทำอะไรเกินเด็กจัง
76	พวกโจมตีสุดลิ่มทิ่มประตูแบบนี้ ยิ่งน่าสงสัยกว่าครับ เป็นหมาป่ามาก่อหวอดหรือเปล่า?
77	เสื้อแดงมีแต่พวกถูกซื้อสมองไม่ค่อยมีคอยแต่ลอกเขาและทำตามคำสั่งเท่านั้นจะ-เขาบอกเผาก็เผาเขาบอกขโมยก็ขโมย-น่าสงสารนะ
78	ความสุขในงาน กับ ความสำเร็จของงาน อะไรมาก่อนกัน ? .... ความสุขภายในตน [Inner Work Life -_-"
79	เค้าว่าการสูญเสียความทรงจำเจ็บกว่า สูญเสียคนที่รักนะเพราะคนที่รักเค้าลืมว่ารักเรามันทรมานแค่ไหน
80	สุดท้ายพวกต่อต้านต้องเป็นคนรับกรรมเมื่อไม่มีความสวยงามของธรรมชาติให้ขายอีกต่อไปถ้าปล่อยให้เป็นเช่นนี้...เศร้า
81	ถึงจะเป็นรสนิยมส่วนตัวก็เถอะนะ แต่ผมอ่านแล้วรู้สึกกลัวแทนผู้หญิงยังไงไม่รู้สิ :roll:
82	เดี๋ยวนะ ไม่สวมหมวกนิรภัย นี่หมายถึงแท็กซี่อย่างเดียวปะ หรืออย่างอื่นด้วย ? =)3(=
83	แต่ก็ไม่ใช่ว่าคุณจะสามารถหนีปัญหาแบบนี้ได้ คุณควรที่จะร่วมหาทางออกกับสมาชิกไม่ใช่ ทำท่าทีงอล แล้วบอกจะไม่ทำแล้ว
84	แหม เข้มขัดสั้นเลยทางนี้ ไม่นึกว่า FZ จะไปอยู่กับกุหลาบ เพราะว่าภาค"ค้างคืน" อยู่กับ Tiga ฮ่า~
85	หงึกๆ อยากรู้เหมือนกัน เจอกระทู้สปอยว่าน่าผิดหวัง ก็ยังอยากผิดหวังด้วยตัวเองอยู่ดี ีรอด้วยคน =3=
86	ถ้าเราไม่ทำอะไร ยอมให้นักการเมืองและนายทุนเลวๆครองเมือง ชะรอยสยามจะเป็นเหมือนเขมร ประชาชนเดือดร้อนกว่านี้แน่
87	อยากลองไปดูจัง แต่ก็กลัว ขนาดเคยเจอมาก็หลายครั้งต่หลายครั้งแล้ว แต่ก็ยังไม่ชินซะที
88	ถ้าจำไม่ผิดมันเป็นชายหาดที่อยู่บนเกาะในมหาสมุทรแอตแลนติกใกล้ฝั่งยุโรปนี่แหละ o_O
89	ผู้ดูแลระบบ คือ คนที่จะต้องคอยดูแลระบบเพื่อแก้ไขปัญหาต่างๆ ไม่ใช่แต่งตั้งมาเพื่อจะมีอำนาจทำอะไรก็ได้ ไม่ทำก็ได้

90	เกินเยียวยาแล้วครับ โกงก็รับได้ขอให้ได้ประโยชน์ 68%, จะประมูล หรือประเคนก็ช่าง ขอให้ได้ใช้ 3 G
91	รอคอย HF ด้วยความสิ้นหวัง ในบรรดานางเอกสามคนมีคุณเธอคนเดียวที่ยังไม่ได้เป็นนางเอกกับเขาซักที OTL
92	อย่าให้นักการเมืองรับผิดชอบโดยการลาออกกนะค่ะ ไม่ได้ประโยชน์หรอกค่ะ ต้องยึดทรัพย์สินคืนมา ฐานให้ประเทศเสียหาย
93	กลัวตอน ผีดูดเลือดอุราโดะมากๆ ถึงขนาดดูจบแล้วเอาไปนอนฝัน ว่าตัวเองเป็นเหยื่อเลย = =
94	นี่ท่าน DarKnight เป็นแนวร่วมกระทืบไอ้แว่นหรอครับ ? ..ไม่รู้น่ะเนี่ย !!! :grin:
95	กะ..กะ...โกง เ-ี่ย ==!!! ไอมังกรอ่ะ สกิลอันติมัน สมมุติอีกฝั่งเกิดทั้งทีม ใช้สกิลนี้ทีเดียวนี่พลิกเลยนะเนี่ย- -
96	สู้ ๆ แก้ตัวได้สำเร็จ แต่ไม่เหมาะกรณีงานระดับโลกครับ ยอมรับผิดและขอโทษ ปชช และผู้เกี่ยวข้อง
97	Mashiro-Iro Symphony ค่ะ ตอนแรกกะจะตามต่อจนจบ แต่เจอสปอยไปที... ป่านนี้ยังดูไม่จบเล้ยยยย....
98	ถ้าคิดแบบนี้ไปเลี้ยงควายดีกว่าไม่ต้องร้องหาประชาธิปไตย ใครบ้างที่เลือกมาแล้วไม่โกง อยากทราบว่ามีสมัยไหนบ้าง
99	เพิ่งได้ดูไปเมื่อคืนครับ แต่ไม่จบต้องไปทำงาน พร้อมกับความหลอนเล็กๆจากหนังเรื่องนี้
100	มีใครได้สั่งซื้อ BD แผ่น 1 มั้ยอะคับ ? อยากถามหน่อยว่ามันมีอะไรมาให้มั่งอะ *-*
101	ไม่ได้ต่างกันเลย ... คือถ้าเปลี่ยนกรอบจากจริยธรรม เปนกรอบ เงิน ทั้ง 2 กรณีล้วนแต่เปนไปเพื่อปกป้องเงิน ทั้งสิ้น
102	พอจะเข้าใจละฮ่ะ เพราะเมื่อก่อนผมก็ชอบตัวละครประมาณ ดาบ2มือใหญ่ๆ เท่ๆ พระเอกเหมือนกัน-////-
103	มันคืออะไรครับ เมื่อคืนเปิดทีวีดู ก็เห็นบอกว่า ห้าม Like ห้ามแชร์ ไม่งั้นทั้งจับทั้งปรับทั้งจำ
104	ตอนมันไปกดหมอนใส่หน้าพี่ชายนะ อื้อหือ แทบจะปรี่ไปตบมัน รู้สึกแม่มันตบเบาไป เป็นเราเราจะเอาส้นสูงตบ ว้าย...
105	ความรู้สึกเหมือนมันจะสารภาพรักยังไงไม่รู้ (อิ๊....ขนลุก) อย่างบอกไม่ถูก :confused:
106	แล้วตรูเป็นซีเหนี่ยแต่เมื่อไหร่เนี่ย..ท่าจะปั้มน่าดู (วันกี่มู้แว้..ไม่ได้นับ)
107	คนไทยน่าจะแสดงสปิริตเหมือนชาติหลายชาติ การรักษาจริยธรรมของชาติและสังคมถือเป็นหน้าที่ของคนรักชาติ อย่างญี่ปุ่น
108	รด.ปี 3 น่ะอาบไม่ได้เลย ต่อให้ลุยมาเละแค่ไหนก็เถอะ ถ้าจะอาบก็อาบวันสุดท้ายแล้วก็เปลี่ยนชุด
109	น้องสาวก็ขายทิ้งไปแล้ว(เพื่อซื้อIS) ส่วนแผ่น 5 พรีเมี่ยม ก็รอในงานหนังสือ (มีเหลือทีเถอะ TwT)

110	เป็นข้าราชการแต่ร่วมกันทำผิดกฎหมายปล่อยให้มีการบุกรุกทำลายป่า งวดนี้นางลำบากแน่ ยิ่งลักษณ์ ทราบแล้วเปลี่ยน
111	มันเงิบซะจนผมไม่คิดจะยุ่งเกี่ยว ถึงจะบอกว่าเป็นเหตุผลทางเนื้อเรื่องก็ตามเลยล่ะครับ
112	ทุกคนต่างเป็นกำลังใจเมื่อนักเลงแถวบ้านท่านasแสดงความห่วงใยอย่างจริงจังไว้แล้ว
113	สันดานโจรถึงอย่างไรมันก็ไม่ยอมรับหรอกว่ามันเป็นโจร มันทุเรศตรงที่ผู้มีหน้าที่เกี่ยวข้องทำตัวเป็นโจรเสียด้วยซิ
114	คิดถึงเหมือนกันแหะสมัยเด็กๆเสาร์-อาทิตย์คือสวรรค์ในการดูการ์ตูน ช่อง9การ์ตูนช่วงฉายเยอะมาก
115	เคยทนกลั้นใจไม่ดู ปล่อยให้มันฉายไปสัก 3-4 ตอนก่อนน่ะค่ะแล้วค่อยดูรวด /แล้วก็มาทรมานรออีกรอบ
116	แล้วถ้าตำรวจทีารับมอบนโยบายเหี้ยๆแบบนี้แล้วเอามาใช้จริงก็กลับไปเอาผ้าถุงเมียมานุ่งเป็นเครื่องแบบก็แล้วกัน.
117	เอ่อ...ข้าพเจ้ายังไม่นอนนะ คนอื่นนอนกันหมดแล้ว เหลือข้าพเจ้าคนเดียว ข้าพเจ้าหลอนนะ
118	ขอโทษนะครับ ทำไมตอนที่เขาบอกว่ายังไม่รู้อีกหรอ เสียงนางโง่มันมาจากไหนหว่า *-*
119	เอากันสักตั้งสิ แบนไม่บริโภคสินค้าที่มาโฆษณารายการของมันซะ ดูสิว่าสปอนเซอร์พวกนี้ยังจะหลับหูหลับตาอยู่อีกไหม
120	หึหึ เกินเดือน2ปีหน้านะ SAO DMG ยังไม่ซึมซับเลย ฮ่า~ เป็นที่แน่นอนแล้วนะ หมดประกาศแล้วสินะ
121	ความจริงมี่จังหลังจากถอดชุด แล้วต้องจับกดจิซาโตะ จากนั้นจึงจบ Harem route นี่มันเป็น OVA...
122	คือว่าผมไปลง window7 pro มาอะครับแล้วทีนี้คือพอลงเกม อะครับ พวก prototype 2 อะไรงี้อะครับ มันเล่นไม่ได้อะ
123	จะบอกว่าน่ากลัวก็ไม่เชิง ตื่นเต้นลุ้นกับนางเอกมากกว่า แต่ก็ทำเอานอนไม่หลับอยู่นะ
124	โรงเรียนอะไรครับเนี่ย จัดหลักสูตรให้เด็กประถม(?)มาทำแบบสอบถามแล้ว :confused:
125	ผมหวังอย่างเดียวแค่ว่าให้เนื้อเรื่องมันเรื่อยๆเปื่อยๆแบบนี้ อย่ามีอะไรปวดตับก็พอ... อยากดูอะไรสบายๆมั่ง = ="
126	เอาSAOไปไม่ว่า แต่ให้มันจบก่อนเถอะ ตั้งใจจะซื้อแผ่นท่านอยู่แล้วล่ะครับ(มันอดไม่ได้หรอก)
127	จริงครับ ผมเป็นคนนึงที่เคยแชร์ข่าวผิดเมื่อนานมาแล้ว(ผ่านมา 1 ปีแล้ว) ตอนนั้นโดนด่าเละครับ
128	อย่าว่าแต่นักการเมืองเลยครับ ขนาดนักข่าวเจ้าของสื่อบางสื่อ ยังมีคนให้ความช่วยเหลือไม่ให้ติดคุกอย่างง่ายๆ
129	The Others ครับ ขนาดตอนยังไม่หักมุมผมดู(คนเดียวปิดไฟตอนกลางคืน) ขนลุกพรึ่บๆๆเลย

130	ต้องไปถามคนบางคนที่พูดว่ามีวันนี้ เพราะพี่ให้นะ..ว่าจะกล้าจับหรือเปล่า!!!!!!
131	ใส่ disk แล้วโดนเตือนก็ไม่เป็นไรครับ เพราะข้อมูลใน disk ลบไม่ได้ // อีกวิธีคือปิด anti ก่อนเล่นเกมส์ครับ ^^
132	เริดนะฮ๊าฟ ท่านด็อกสอบได้เกรด A วิชากฎหมายอิสลามของ ม.รามคำแหง สาขาดาวอังคารรึเปล่าอ่ะ
133	ใครเจอโก๊ะตี๋ช่วยถามเตือนสติให้ทีว่า "ยังจำได้หรือเปล่าเพราะอะไรถึงแยกวงจากเอ็ดดี้ผีน่ารัก
134	แล้วเห็นเรื่องจะบัญญัติศัพธ์ใหม่ตามภาษาวัยรุ่น วันรุ่นก็เสือกไปด่าเขาอีก ทั้งๆทีตัวเองเป็นคนสร้างมาแท้ๆ
135	ปกติผมจะรอให้เมะออกอากาศจบก่อนค่อยเอามาดูรวดเดียวเพราะไม่อยากให้อารมณ์ขาดช่วง
136	ไม่ใช่อะไรหรอก ผมแค่อยากรู้เหตุของแรงโทสะที่เกิดขึ้นในใจของเด็กคนนั้นมากกว่า
137	วางตัวกลางๆ ไม่ต้องแก้ตัว งานระดับโลก มีเวลาตั้งหลายปี มัวทำไรอยู่ ช้างตายทั้งตัวรับรองเหม็นเน่าทั้งป่าครับ
138	ใครแน่เริ่มก่อน. แต่จริงอยากให้สรยุทธกล้าเปิดความจริง ผมว่าเบื้องหลังต้องมีทีเด็ดแน่ๆๆ
139	เศร้าครับ ข่าวแบบนี้ แต่เท่าที่ผมเจอกับคนจีนในไทย ผมว่าทุกคนเป็นคนมีน้ำใจกันหมดเลยนะครับ
140	คนไทยจนเอาจนเอา แต่รัฐบาลรวยเอารวยเอา ป ป ช ควรต่อต้านการคอร์รัปชันของนักการเมืองไทย ก่อนทีจะสายเกินไป
141	เออ...นี้มันก็ตี4ครึ่งแล้วนะครับ ทำไมต้องเอาของน่ากลัวแบบนี้มาให้ดูด้วยเนี่ย
142	อยากรู้ว่า พวกงาน event อะไรบ้าง ที่มีบูธ หมุนล้อเสี่ยงโชคอะคับบ :confused:
143	สมเพชไอ้เหลี่ยม ต้องพึ่งมือ พึ่งตีน คนอื่น ตัวมันเอง ไม่มีปัญญา ทำเชี่ยอะไร ไอ้พวกนี้ก็ทำมาหาแดกไม่เป็น
144	เมื่อคืนกินเหล้าปลอมป่าววะเนี่ย กินแล้วไม่เมาเรยย เอาไว้กินกันใหม่ละกันนะเพื่อนๆ ...เย่
145	ซื้อ CPU มือสองไปเมื่อ 3 เดือนที่แล้ว ตอนนี้ราคาตกไปแล้ว 600 บาท เสียหายไปเดือนละ 200 T_T
146	อิฉันประสบกับปัญหาเล่นเกมส์ออนไลน์แล้วมักจะโดนเกรียนด่าให้ไปเล่น Mario ถ้าเล่นแล้วกาก ขนาดนี้ !!!!!:(
147	ฉากน่ากลัวก็น่ากลัวโดยที่ไม่ต้องใส่เสียงดังๆ ให้คนดูสะดุ้งเหมือนเรื่องอื่นๆ
148	จริงๆ ฟิวเนรุ ไม่ได้อยู่ในเหตุการณ์ แล้วทำไมมันเข้าใจมากกว่าไอ้สองคนนี้ ฟระ
149	ถ้าไม่ดีจริงคนกว่าครึ่งประเทศไม่เลือกเป็นนายกหรอก รู้จักป่ะประชาธิปไตยอ่ะหรือรู้จักแต่ประชาธิปัตย์get ป่ะ
150	อ่านข่าวนี้แล้วตลก แสดงว่า Apple ใช้ เลนส์ ไม่ดี ทำให้เกิดอาการขอบม่วงเวลาถ่ายย้อนแสง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

151	ไอพวกที่โกงไปมันคิดว่ามีอายุอยู่บนโลกนี้เป็นร้อยๆปีหรอ สุดท้ายปลายทางชีวิตก็กลับมาเป็น 0
152	ไอ้หน้าไม่อาย สัสหมาเหมือนพวกมึง ทำให้คนดูข่าวต้องเหม็นหน้ามัน ทั้งที่อยากดูข่าวคนอื่น สัสเอ้ย สัสสสส
153	เวรกรรมมีจริง....เงินที่มันได้ไป..มันไม่ได้ใช่หรอก...เอาไปรักษา...มะเร็ง...
154	ไอ้คิริโตะมันไปอยู่กับพวกนี้ได้ไง มันแค่ชอบเล่นเกม แต่ยังไม่ถึงขั้นเพ้อนะ
155	The Other .... ทำให้ผมเกลียดพันทิพจนเกือบกระโดดเอาไม้ฟาดหน้าจอเมื่อเจอพวกมือบอน มาโพสต์ตอนจบแกลังคนเล่น
156	เป็นอีกดราม่าที่หาใจความไม่ได้แล้วหน้าก็ยาวมาถึง 7 หน้าผมละเงิบจริงๆให้ตายสิ Ha Ha Ha
157	ชีวิตตอนนี้เหมือนโดนสตอล์กเลยค่ะ เหมือนแสงอิสระเริ่มลางๆหาย /ช่วงแรกไม่ใช่แบบนี้เลย TT
158	สรุปแล้วไอ้พวกเสื้อแดง เพ้อเจ้อ อย่าไปฟังมัน พวกนี้ แม่งตายแล้วเกิดใหม่ สมองมันก็เล็กกว่า หิ มดหน่อย
159	รอข้างนอกไม่ไหว ห้องน้ำคนเยอะ หากับข้าวซื้อลำบาก จนสุดท้ายไปนั่งหลับในรถ...
160	ไม่ใช่ว่าทีเรามีอยู่ชีวิตอยู่ได้เพราะโลก(เตี้ย)แบน(ซืน)หรอกเหรอ :rolleyes:
161	หยุดอ่านข่าวสักพัก แบบมาร์ธา สจ๊วต ไปสู้คดีให้เสร็จ แล้วกลับมาทำใหม่ไม่ได้หรือไง รวยจะล้นฟ้ากันอยู่แล้ว
162	วิธีที่ดีที่สุดคือไม่ดูรายการข่าวของเขาและเมื่อเรดติ้งต่ำลงผู้สนับสนุนก็ถอนไปเรื่อยๆ
163	แต่บางครั้งการสูญเสียความทรงจำอาจจะทำให้เราเสียช่วงที่เคยมีความสุขกับใครซักคนก้ด้ายนะ
164	รัฐบาลชุดนี้เค้าให้พวกคุณรู้แค่นี้จริงๆเหรอ เค้าเห็นพวกคุณเป็นอะไรกัน..... อยากให้หาข้อมูลให้เยอะๆ
165	เคย เพราะกลัวที่จะต้องเสียน้ำตา เพราะคำถามที่ถามไปแร้วไม่ได้คำตอบที่หวังไว้
166	โฮ้ .. โฮ ไม่ได้เข้า Tirkx มาวันกว่า..วงซับนรกไปต้อง 4 ตอนแล้วหรอนี้ !!!!
167	แล้วมันประท้วงหาสากไรอีกล่ะ ช่วยเอาน้อยหน่า โยนไปให้ สักลูกสองลูก มันจะได้กลับนรกถูก อยู่แล้วรกแผ่นดิน
168	ขอขอบคุณแถบ sample เราจะไม่มีวันลืมคุณงามความดีครั้งนี้ของนายเลย เจ้าแถบ sample เอ๋ย
169	ผมเป็น mod เหมือนกันที่บอร์ดอื่น ผมไม่กล้าสั่งสอนใครหรอก เพราะผมก็ไม่ได้ perfect อะไร
170	เงินหมื่น หกพันล้าน นี่แบ้งค์ ชาติไม่รู้เรื่องเลยหรือ ตั้งงบไปเที่ยวฮองกง มากกว่า เอาเรื่องนี้มาอ้าง
171	ซับนรกแบบสบาย ๆ น่ะครับ .. ไม่ต้องเอาซับนรกแบบติดระเบิด(ระเบิดตับ)มาด้วยน่ะ
172	เท่าที่อ่านมาจากเวปนอกว่าอาจมีต่อภาค 2 มั่ง ไม่แน่นใจเพราะอ่านมานานมากละ

173	เบื่อหน้าฉิบหาย ไม่ดูมานานแระคำช่อง 3 ดูไทยพีบีเอส ดูเนชั่น ดูมวยช่อง 7 วันอาทิตย์ ดูช่อง 9 บางรายการ
174	เป็นโรคสายบ่อยเหมือนกัน อีกโรคคือ ขยันกับเรื่องไม่เป็นเรื่อง... (นอกนั้นขี้เกียจหมด)
175	ถ้าปรับให้ไร้ความรู้สึกไม่ก็อาเรียถูกฆ่าต่อหน้าก็คงจะเข้าDark sidenกล้าควักหัวใจละครับ
176	ยังไม่เคยใช้ลองจัดแข่งอะไรเลยถ้าเกิดผิดพลาดแล้วทำไง ทำไมชอบแถจังวะพวกนักการเมือง Thailand standard.
177	ถ้าให้ดูสด ๆ คงไม่กล้าดูแน่เลย มีแต่ภาพสยอง ๆ ทั้งนั้น แต่ดูแบบนี้ดูได้ค่ะ
178	ไม่มีทางอะ CPU มันจะทำให้น้ำมันเดือดได้ขนาดนั้นเลยเหรอ น้ำมันเยอะนะนั่น!!
179	เอาจริงๆปัญหามันจะไม่เกิดทั้งที่รู้อยู่ว่าเรื่องนี้มีคนทำสปอยเปนประจำอยู่แล้วจะไปตั้งอีกสปอยอีกทำมั้ย
180	โดยส่วนตัว ถูกใจกับการ LC ครั้งนี้ เพราะ SAO ของข้าพเจ้าจบไปตั้งแต่ตอน 14 ละ (ฮ่า~)
181	คนไหนน่ารัก เด่น มีอันเป็นอาหารให้เบต้าไปหมด ถ้าไม่แกร่งจริงอยู่ในเรื่องนี้ได้ไม่นาน
182	งี่เง่า คนที่เค้าเชี่อ เค้าดูก็มีกัน มาด่าเค้าไม่ดูตัวเองเลย เจอเข้ากะตัวเองซักวัน มึงจะพูดไม่ออก
183	เหมือนคนไม่รู้จักกัน เหมือนเรื่องราวที่เกิดขึ้นฉันแค่ฝันไป อ๊าก ฝันร้าย TT
184	ถามว่า คำว่า รับผิดชอบต่อสังคม กับ คำว่า ละอายต่อการกระทำผิด มีบ้างไหม...
185	ไอ้หน้าไม่อาย สัสหมาเหมือนพวกมึง ทำให้คนดูข่าวต้องเหม็นหน้ามัน ทั้งที่อยากดูข่าวคนอื่น สัสเอ้ย สัสสสส
186	ผมเชื่อว่าถึงเราปิดประเทศเราก็ไม่อดตาย แต่ถ้านักการเมืองโกงกินต่อไปเราอดตายแน่นอน
187	บางครั้ง โฆษณาตรงๆนี่ล่ะ เวลาที่เราต้องทนดูทนฟัง มันนานมากกว่าช่วงเวลาของข่าวเสียอีก
188	อ่านข่าวก็ชอบชี้นำ ชอบบิ๊วเรื่องเล็กให้เป็นเรื่องใหญ่ บ้าอำนาจ ชอบเบรคลูกน้อง เอะอะไรก็เปิดกองทุน
189	อ๊อกกกกกกกกกกกกกกกกกกกกกกกกกกกกกก ดีนะที่ดูภาพนิ่ง ไม่ใช่ภาพเคลื่อนไหว T.T
190	แล้วในเนื้อเรื่อง ALO นี่ พวกโตะกับสุกุนี่ แบ่งเวลาใช้ชีวิตกันยังไงอะ...
191	muv-luv te ครับ เพราะเป็นเรื่องที่อวยที่สุด เดินเรื่องไม่ได้อย่างที่หวังเลย รู้สึกเบื่อๆ + เผาเกรียม
192	เข้า workshop ด้วยภาสาอังกฤษที่อ่อนแอ คุนแม่ปวดท้องใกล้คลอด ห้ามผู้ชายเข้าใกล้..
193	เห็นแล้วก็นึกสงสารเฮียคริส ภาคนี้คนอื่นได้พาร์ทเนอร์เป็นสาวๆหมด ยกเว้นเฮียแกคนเดียว

194	รู้ว่าประเทศไทยนักการฉ้อโกงข้าราชการคอรัปชั่นแต่นี่มันเหลือเกินน่าสงสารๆพี่น้องเราคนเสียภาษีตาดำๆ
195	คนขี้เมา...พูดให้บ้านเมืองเกิดความแตกแยก...สักแต่ขอให้ได้พูดเพื่อความสะใจ
196	เรื่องมีอยู่ว่า หมา มันมาจากไหนไม่รู้มันชอบมากัด รองเท้าที่วางไว้หน้าบ้า
197	คงถึงเวลาต้องคุยกันแล้วล่ะมั้ง คุยให้เข้าใจกันอะ ไม่งั้นอยู่ไปนานๆ ยิ่งรับไม่ได้ยิ่งทรมานกันเปล่าๆ
198	ท่านด็อกสอบได้เกรด A วิชากฎหมายอิสลามของ ม.รามคำแหง สาขาดาวอังคารรึเปล่าอ่ะ หุหุ
199	แค่ปัจจุบันนี้ใครเข้าgoogle mapดูแผนที่ดาวเทียมจะรู้ว่าประเทศไทยป่าเหลือน้อยสุดแล้ว
200	คนเราความผิดของตัวเองมองไม่เห็น ไม่สนใจ ไม่รับผิด ใครจะทำไม ตรูมีตังค์ซะอย่างเฮ้อออ! ประเทศชาติ
201	ดูแต่ Sky rider ช่วงรวมเรื่องผี ดูจบแล้วไม่กล้าอยู่คนเดียวไปหลายวัน = ="
202	Case study เราเอง.. มันซับซ้อนไปอะ.. บางครั้งเราเองยังไม่เข้าใจเลยย 555+
203	ถ้าลูกหลานเปิดซ่อง3เจอตัวเหี้ยนี่ ต้องรีบอธิบายว่าไม่ดีอย่าดู ตัวเสนียดสังคมรังเกียจมันตัวโคตรโกง
204	LC จัดมา เก็บแน่เรื่องนี้ แต่ขอดูภาพตอนตกปลาใน BD ก่อนว่าจะแก้ภาพนั้นรึเปล่า ..
205	ขอโทษที่รักษาชีวิตแกไม่ได้ หวังว่าเหมียวจะได้เกิดใหม่ในสภาพที่สมบูรณ์กว่านี้นะ RIP
206	มันเวียนกันมาจำนำกี่รอบก็ไม่รู้ ขอให้ยุบ อคส. ซะ โดยเฉพาะ หัวคะแนน เพื่อไทย + เกษตรกร อยากได้เงิน
207	ซอร์ด มาแต่แรก เพราะเจอจารย์เฮี๊ยบสุดๆ สมัยประถม อ่านผิดเจอไม้เรียว ฮ่า~
208	อยากถามว่า ทุกวันนี้เงินเดือนพวกเมิง ไอ้แม้วมันจ่ายให้เหรอ พวกส้งตี....
209	ที่ผมไม่เข้าใจก็คือลบของคนอื่นแล้วมาโพสต์สปอยเองโดยที่ไม่ไปโพสต์ในโพสต์หลักให้มันเหมือนกับคนอื่น?!
210	พ่อใช้วอลเปเปอร์เคโรโระ ไม่พอ ใช้ตีมเคโรโระด้วยค่ะ .. /พ่อเราเป็นคุเลยสินะ = =;
211	อ้าว มีต่อด้วย เสียดายไม่ได้อยู่ดูต้องรีบกลับมาก่อน เดี๋ยวไม่ทันอ่านหนังสือสอบ ^^
212	สื่อทำหน้าตรวจสอบคนอื่นแต่พอกับเจอกับตัวเองกับจะไม่ยอมรับการตรวจสอบยังเรียกตัวเองว่า สื่อ อีกหรอ
213	ได้แค่เกรด แต่ความรู้ชุ้ยๆ ไม่ได้ทำผิดกฎหมาย แต่ทำในสิ่งที่กฎหมายห้าม !!
214	เฮ่ยผู้ชายหรอกหรอ ฟังจากมือถือเพื่อนบ่อยๆแต่ไม่แน่ใจเสียงมันสับสน - -
215	สำหรับคนที่บอกว่าผมไปลบกระทู้เขาทั้ง ๆ ที่เขามาตั้งก่อนก็ช่วยพิจารณาข้อมูลด้วยครับ ผมขี้เกียจพูดละ
216	ชีวิตผมจะเจริญขึ้นมั้ย??? ถ้าไม่...ผมใช้แบบเดิมดีกว่า สะดวกกับชีวิตผมดีครับ...
217	โอ๊ย....เจ็บจากตอนแรกไปหยก ๆ มาตอน 2 ...ไม่มีคำบรรยาย แต่สภาพตัวเป็นยังงี้ => TOT

218	เลิกดูรายการของคนโกงกินแล้วครับ ทั้ง 3 รายการเลย หลงคิดว่าเป็นคนดี..ที่แท้ก็อมเงินค่าสปอนเซอร์..
219	โอ้ววววววววววววววววววววววววววว!!!!!!!!!!~ ขนาดภาพนิ่งนะเนี้ย จะเป็นลม~
220	ตอน10มาแว้ววววววววว ว่าแต่ไหงไฟล์มันใหญ่ขึ้นได้นิ หรือผมคิดไปเอง =w=a
221	ใครไม่เคลียร์เรื่องไหนอีกก็ถามมาดูละกันครับ ผมขอเลิกยุ่งกับปัญหา ของ Anime เรื่องนี้ตอนนี้เลยละกัน
222	ทำไมการ์ตูนฝรั่งหลายๆเรื่อง ถึงไม่เห็นครึ่งบนของตัวละครที่เป็นมนุษย์ผู้ใหญ่ค่ะ
223	โอ๊ย....เจ็บจากตอนแรกไปหยก ๆ มาตอน 2 ...ไม่มีคำบรรยาย แต่สภาพตัวเป็นยังงี้ => TOT
224	อันนี้มันงี่เง่าไปครับ โตๆ กันแล้ว ช่วยกันคิดหาทางออกทีพอใจทั้งสองฝ่ายจะเป็นเรื่องที่ดีกว่าครับ
225	ดูจบ 4 ทุ่ม นั่งรถเมล์มาต่อรถที่อนุสาวรีย์ ยังตัวสั่นๆ กลัวแสนจะกลัว
226	subtitle มันผิดหรือเปล่า? เพราะ ที่พิธีกรพูด กับ ภาพมันยังตรงกันนะ :?:
227	คุณคิดหรือว่าพวกนั้นจะยอมปรับทัศนคติ... ถ้าเขายอมรับได้เขาคงไม่ออกมาโวยวายหรอกครับ หรือไม่จริง...
228	ขับรถกลับบ้าน อย่างกับขับเรือกลับบ้าน ดีป้ายทะเบียนไม่หาย ละก็น้ำไม่เข้ารถด้วย
229	อากิโกะของหนูบทก็ยังตัวประกอบเหมือนเดิม มันไม่ตรงตามชื่อเรื่องนะเนี่ย ม่าย TT^TT
230	ไอ้นี่มันชอบอ่านข่าววัวหลุดจากรถ. หลุดจากโรงฆ่าสัตว์. สุดท้ายลืมอ่านข่าวความเป็นสัตย์ของตัวเอง
231	อึม ปกติคงไม่กล้าดู มาดูหนังกลางแปลงอย่างนี้ค่อยรู้สึกไม่น่ากลัวหน่อย
232	ขับไปชนมาอ่ะ"ฝากระโปรงหน้าเปิด""""ก้เลยซ่อมสีใหม่ทั้งคันด้วยเลย""""หุหุ
233	ประเด็นผมคือถ้ามีคนตั้งเป็นประจำอยู่แล้วจะตั้งไปเพิ่มหาสวรรค์หรืออย่างไร ไปตั้งเรื่องอื่นเอาก็ได้
234	งวดหน้าลองเป็น Hall อย่าง BCC หรือ MCC ก็ดีนะครับ Impact มีรถรับ-ส่ง นิน่า -*-
235	แต่ดูจากท่าทางวันนี้ผมคงจะโดนโหวตให้ออกจากบ้านกระต่าย เอ้ยโดนแขวนคอแล้วสิลาก่อน
236	อ้าว..นึกว่าตำรวจของประชาชน..เป็นตำรวจของนักโทษชายหรอคับ..แล้วแดกเงินเดือนใครกันล่ะ..ถุยส์..!!
237	นอนคนเดียวด้วย โครตน่ากลัว อ่ะ มีคัยดูเปนเพื่อนสักคนน้าจะดีขึ้น >< !
238	คอมพิวเตอร์ -> ค็อมพิ้วเต้อร์ มันไม่ไหวนะไอ้ตัวนี้ แปลกๆไงก็ไม่รู้ :(
239	ดูความหน้าด้านสรยุทธ์ ยังนั่งหน้าหนาอยู่ทุกเช้า คนนั่งข้างๆ ไม่รู้สึกเลยเหรอว่านั่นมันไอ้ขี้โกง
240	น้อสาวผมสั้นเราไม่ได้หาเจอแล้วถูกใจง่ายๆนะ เพราะฉะนั้นออกบททีตูจะเฮให้ลั่น...

241	เมื่อ User เกิดปัญหาคุณก็จะต้องช่วยแก้ไข ช่วยดูแลให้ระบบสามารถดำเนินงานต่อไปได้
242	อยากจะกลับไปยุคที่ ไปติดต่อกับทางราชการแล้วโดนพวกมันตะคอก ดุด่าดูถูก ใช้เวลาตามใจพวกมันอีกหรือ
243	ถึงจะเจ็บยังไงแต่หัวจัยคนเรามีแค่ดวงเดียวเมื่อคิดจะรักอย่าไปกลัวเจ็บ
244	ขอทำมั่งดิ..อยากรู้คราวนี้จะมาแนวไหน (ครั้งก่อนระลึกชาติเอาเรื่องเลย)
245	เมื่อไหร่ของที่เคยได้ถูกแย่งไปก็ออกมาโวยวายเพราะตนเสียผลประโยชน์ หรือสั้นๆ " เห็นแก่ตัว " ไงครับ
246	ผมคิดว่าเราไม่ควรดราม่า เพราะทุกอย่างที่ทำ ทำเพื่อความสนุกไม่ได้ทำเพื่อดราม่า
247	กรี๊ดดดดดดดดดดดดดดดดดดดดดดดด ไม่ยอม จบงี้ได้ไง อิน้านเอาลูกปายได้ไง ไม่ยอม ๆ ๆ ๆ
248	อย่ามาเถียง ถนนประชาสโมสรหน้า บขส แล้วก็เส้นหน้าเมืองอีก เห็นแต่วิญญาณเส้นน่ะสิ ลางเลือนมากกก
249	the others เป้นหนังผีเรื่องแรก เรื่องเดียวในชีวิตที่เรากลัวมากกกกกก
250	นี่มัน... รู้สึกว่าจะเป็น... เขาเรียกว่าอะไร... อีลูเนียม.....สินะ
251	สามัญสำนึก เป็นเรื่องที่เรียนรู้หรือสอนกันไม่ได้...มันต้องเกิดจากอินไซด์ของตัวตนเองเท่านั้น....
252	สนใจ Ryougi ง้าบบ ผมไปbig fest พรุ้งนี้เหมือนกัน วันเสาร์ใช่ปะ งั้นนัดรับเลย
253	คนอวดผีไม่น่าย้ายช่องเลยช่อง7นั้นเละดีแล้ว ภาพชัดดี พอย้ายช่องดูไม่ค่อยชัดเลย
254	เหี้ยยังไงก็เหี้ยอย่างนั้น สำหรับไอเหลิม เก่งจังเลย ลูกแม่งก็ยังไม่ใช้กรรม เสือกได้ตำแหน่งอีก
255	ไม่กล้าเสี่ยงกับโรคที่ตามมา ปล่อยเลยอ่ะ ถ้าจำเป็นจริงๆ เต็มที 1 วัน
256	นี่มัน... รู้สึกว่าจะเป็น... เขาเรียกว่าอะไร... อีลูเนียม.....สินะ
257	ก็ไม่มีอะไรมากหรอก ก็แค่อยากรู้ว่ามีใครสังเกตุความต่างของ op Ebiten แต่ละตอนได้บ้างเท่านั้นเอง
258	โคนันมันเตี้ยเปิดประตูไม่ถึง แต่พอมาภาคหลังๆมันสูงขึ้น(เปิดประตูเองได้)*O*!
259	ง่ะ เพลงที่ของวงนี้ก็มีในเครื่องโดยเฉพาะเพลงของDarker Than Black ยุบซะงั้น= =
260	หลายคนโลกสวยคิดว่าจ่ายๆไปเถอะภาษีหน้าที่ผู้มีรายได้ ถามหน่อยเวลาลูกแบมือขอตังถามมะเอาไปทำอะไร
261	เห็นเป็นแผ่นใหญ่ๆมาเลย ใช้อย่างไรครับ ไม่กล้ารุนแรง เดียวพังนะครับ
262	ดีฟคิสแบบไหน จะได้เป็นผู้เล่นสีส้มหรือสีแดงละเนี่ย มันเข้าข่ายPKนะนิ
263	เลิกดูข่าวสรยุทธมาหลายปีแล้ว เพราะแกซุกไซ้ข่าว เจ๊าะแจ๊ะ ดูเหมือนจะละเอียดทุกเม็ด แต่ก็ไม่ชัวร์
264	กองกำลังท่าจะเยอะแฮะ รอดูต่อไปก่อนแล้วกัน ว่าจะ "ช้ำ" มากแค่ไหน/เผ่นนนนน !
265	ทำใจล่วงหน้าว่าถ้าไม่ดูซับอังกฤษคงอีกนานเป็นชาติครับ สารภาพบาปแล้วนะเนี่ย

266	ค่าเหี้ยอะไรของมันตั้ง7หลัก ชื่อรายการที่ต้องจ่ายแลดูไม่น่าเยอะนะ แต่แม่งพอดูด้านขวาจะเป็นลม
267	นางเอกโหดมาก ทำการผ่าตัดด้วยตัวเอง (ฉากนั้นต้องปิดตากันเลยทีเดียว)
268	รับเงินสินบนเอาเงินมาทำบุญ เอามาช่วยหมาแมวคนแก่ แบบนี้ใจบุญรึเปล่า?
269	ผมไม่รู้หรอกนะว่ามันจะเป็นเรื่องอะไร แต่การหนีปัญหามันไม่ใช่เรื่องที่ผู้ดูแลระบบจะทำหรอกนะครับ
270	บอกตามตรง ผมเห็นชื่อมู้แล้วผมก็เดาได้ทันทีว่าใครตั้ง แถมยังเดาถูกซะด้วย 555
271	ตอนนี้ ฉันมีปัญหา เข้า xat ไม่ได้ค่ะ ถ้ามีข้อสงสัย post ในกระทู้ได้เลยค่ะ T T
272	เป็นธรรมดาเมื่อขึ้นสูงตอนลงก็เจ็บหนักกรรมไม่เคยเว้นใครเห็นอย่างนี้สร้างบุญเว้นบาปกันเถอะครับ
273	ตกใจช่องสุดท้ายเฟร้ยยยยยย เหยดดดดดดดดด!!!!! :shock: :shock: :shock:
274	.เอ่อ..มีใครสงสัยเหมือนผมไหม ? คือ อายาเสะ หน้าคล้าย คิริโนะ ว่าไหม
275	ช่ายเวลามีข่าวเกี่ยวกับแบงค์ใดที่ๆๆไม่ได้ซื้อเวลาจะกระทืบให้จมดินเลยจนแบงค์นั้นต้องซื้อเวลามัน
276	แต่พอมาดูk-on! =w=a ดูได้แค่2ตอนก้ขอพี่ชายซื้อกีต้าร์ตัวแรกในชีวิตแล้ว5555
277	SAO ออกหลังหมูบินแผ่น 1 สัก 2-3 เดือนนะครับ ไม่งั้นเก็บตังค์ไม่ทันแน่นอน = ="
278	เป็นช่วงเวลาที่หิวจัดๆและกะจะกินของที่เก็บไว้แต่ดันโดนใครไม่รู้มาแอบกิน- -*ปริ๊ดแตกนะบัดนาว
279	ช่วงแรก ๆ แอบปิดหน้าน้องหลอนครึ่งหลอน หน้าฝั่งสวยนี่สวยใช้ได้เลยนะ
280	ที่ผ่านมา 50 ปี ไทยนักการเมืองโกงกิน แต่พม่าแช่เข็ง จะเอาแบบไหนล่ะ
281	อย่าหลงประเด็นนะครับ ไม่มีใครเหยียบใครนะครับ ภาษาที่ใช้ก็ไม่ได้หยาบคาย แค่แรงเพื่อให้มันรู้สึก
282	ขอบคุณ app store online ที่ช่วยให้ผมไม่ต้องไปบ้าซื้อตามกติกา ปญอ. ของ DTAC
283	แต่ตอนนี้ไม่รู้ว่าจะยิ้มหัวเราะหรือร้องไห้ดี...ทำอะไรไม่ถูกแล้วอึดอัดสุดๆค่ะ
284	เป็นเรื่องของรัฐบาล แต่เงิน"ภาษี"ของประชาชนมาใช้หนี้ แน่จริงเอาเงินพงศ์เทพใช้สิจะไม่ว่ากัน
285	ไม่น่ากลัวอย่างที่คิด ดูไปได้เรื่อยๆ แต่มันน่าสนใจหลังจากดูหนังจบ
286	มี LN Another 2 เล่มจบ ลงได้ไหมครับ (แค่สองเล่ม คงไม่ได้มั้ง แฮะๆ)
287	มันไม่เคยเรียนเลยแล้วแถมจะเป้นด้อกเตอร์มันรู้ไหมว่าคนกว่าจะได้ด๊อกเตอร์น่ะต้องเรียนหนักแค่ไหน
288	ถ้าบนโลกมนุษย์มันยังไม่ได้ชดใช้กรรม เชื่อเถอะ! มันได้ชดใช้กรรมในนรกแน่ๆ!!!
289	วันนี้เนตเป็นอะไรน้อ 5นาทีหลุด 10นาทีหลุด วันนี้เป็นแบบนี้ทั้งวันเลย (-__-")
290	การปฏิรูปการศึกษาของชาติ มันทำได้แค่หาที่กู้เงินเท่านั้นหรือ พวกมันควรปฏิรูปตัวมันเองก่อน..
291	กรรม ทำใจตั้งนานก่อนจะกดดู กะว่าจะเจอรูปน่าขยะแขยง แบบพวกศพอะไรงี้

292	เข้าใจมั้ยหว่า = =" [แล้วถูกด้วยรึเปล่า ไม่มั่นใจ รอเรปล่างๆละกัน]
293	มันไม่เคยเรียนเลยแล้วแถมจะเป้นด้อกเตอร์มันรู้ไหมว่าคนกว่าจะได้ด๊อกเตอร์น่ะต้องเรียนหนักแค่ไหน
294	นี่มันไอเท็มในตำนาน !!! ขอให้เปลี่ยนได้นะคะ เพิ่งเคยเจอแรร์แบบนี้นี่แหละ
295	วันหลังท่านเหลิมอย่าไปบอกใครว่าจบจากรามอีกนะคับ ผมอายเขานะ เสียชื่อสถาบันหมด
296	งั้น Mod เค้ายอมให้ตั้งแล้ว ก็ตั้งกันอย่างมีมารยาทด้วยแล้วกัน ไม่ใช่สักแต่ตั้งๆๆ ท่วมบอร์ด
297	ดูเดือนก่อน....ฝันร้ายจนถึงตอนนี้T T ตอนที่ดูดีนะ ที่ไม่มีใครอยู่
298	เผ่าที่โตะเลือกนี่ มันเก่งด้านไหนอะคับ ? มีท่าไม้ตายมั่งป่ะ ? *-*
299	ระวังให้ดีนะ ปลาตายตัวเดียว เหม็นไปทั้งเข่ง ครอบครัวข่าว 3 ก็จะพังพินาศ เพราะคนคนเดียวนี่แหละ
300	หรือว่าจะถูก Avast ตัวดีของผมมันแดรกไปซะก่อนหมดแล้ว -*- แคร็กผมก็ถูกนะ...

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ค
ผลงานตีพิมพ์จากวิทยานิพนธ์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Emotion Term Feature Selection for Text from Online Social Network

Supakit Chuetanapinyo
Department of Computer Engineering
Faculty of Engineering,
Chulalongkorn University
Bangkok, Thailand
azu148@hotmail.com

Proadpran Punyabukkana
Department of Computer Engineering
Faculty of Engineering,
Chulalongkorn University
Bangkok, Thailand
Proadpran.P@chula.ac.th

Atiwong Suchato
Department of Computer Engineering
Faculty of Engineering,
Chulalongkorn University
Bangkok, Thailand
Atiwong.S@chula.ac.th

Abstract— **Since social media has become part of our daily lives, it has posted some direct effects on technologies. This includes the classification of emotions from text. Specifically in Thai language, the trend in writing in social media has changed enormously to use informal language in which the spellings are often incorrect. Furthermore, users tend to express feelings through writing by imitating sounds. Apparently, these words are nonexistent in the dictionary. Therefore, it is almost impossible to identify emotions from this new writing style. We propose to improve the classification of emotions from text by adding "owner-reader" matching algorithm to the traditional $\chi 2$ (Chi-Square) supervised term weighting method. In our experiment, we performed our algorithm on 3,390 documents from social network and showed an accuracy of 88.79%, which outperformed the traditional unsupervised and supervised term weighting methods.**

Keywords—Emotion, Term Selection, Supervised Term Weighting, Human Emotion Recognition

I. INTRODUCTION

In Human Emotion Recognition (HER) [1], the features in the algorithms to recognize emotion is similar to those human use to perceive emotion. It is a classification technique on given text according to predefined emotional classes. In English text, Danisman and Alpkocak [1] proposed Vector Space Model (VSM) classification to classify the text and achieved 36% accuracy. For Thai text, Inrak [2] proposed Latent Semantic Analysis (LSA) to reduce Eigenvector matrix dimension. LSA can reduce complexity of matrix from meaningless term. This method offered 50.67% accuracy on Support Vector Machine (SVM) and 62.67% accuracy on Naive Bayes (NB) classification. For Korean language, Cheongjae [3] proposed the emotional salience to identify relationships between words and emotions in order to improve term weighting function. They found 84.24% accuracy on Maximum Entropy (ME) classification.

In social media like Facebook and Twitter, the users often express their feelings and emotions on their walls. The writing styles found are mostly informal and unstructured, especially in Thai language. In addition, there exist some vocabularies not present in the dictionary. To classify emotions from these new words can be quite challenging. Moreover, we believe that there may be some discrepancies of emotions perceived by the owner of that text and by the readers. Therefore, unlike previous research where emotions are labeled by a person or multi-person who initiate the corpus, in this work, we will take into account the votes from both owners of the texts and the readers of that text. In addition, we exploit supervised term weighting method to extract emotional keywords from text and term selection method to reduce the dimension from high-dimensional data. Finally, we evaluate our proposed algorithm with traditional term weighting method.

Feature selection is one of the key steps to detect emotions. Previous research showed the use the stop-words list [1] and part of speech (POS) [2, 3] to filter words that do not indicate emotions. When stop-words lists are utilized, it removes hypothesized text that do not have effects on the meaning of the whole text. Usually, they are the high frequent words or low frequent words found in the training data. With POS filter, the part of speech that does not carry important information is removed. This filter often remains Noun, Verb, Adjective and Adverb. However, both the stop-words list and POS filter have assumption that removal word is meaningless. In our work, we argue that these predominant filtering methods may pose some biases especially when words are not existent in the dictionary. As a result, it could contribute to loss of information. As such, we propose the use of based-statistical method to identify occurrences between words and emotions without pre-filtering stage, combining with supervised term weighting method.

Another important task in emotional keyword extraction is dimension reduction. Main concepts include frequency analysis, sentence structure analysis, linguistic knowledge, and part of speech analysis. We chose to employ frequency analysis in our work.

Yang and Pedersen [5] proposed term selection that estimate co-occurrence frequency between term and class such as Document Frequency (DF), Information Gain (IG), Mutual Information (MI), Chi-Square (CHI) and Term Strength (TS). The result showed that IG and CHI are most effective, followed by DF and TS, then MI. All five term selections were tested on kNN/LLSF classifier. Wang, Khoshgoftaar and Hulse [6] proposed feature ranking and feature subset selection techniques that revealed good classifying power.

The term weighting method can be classified into two categories, unsupervised term weighting and supervised term weighting [7]. The common unsupervised term weighting has binary term frequency (*tf*) [8], and term frequency-inverse document frequency (*tf.idf*) [8], [9]. For supervised term weighting, Lan et al. [7] proposed term frequency-relevance frequency weighting (*tf.rf*) and compared with traditional term weighting on various corpus. This term weighting achieved 68% accuracy on SVM classification. Debole and Sebastiani [10] proposed term weighting by replacing the *idf* factor with the metrics that have been used for term selection [5] process such as information gain, Chi-Square (χ^2), and gain ratio. The result of *tf.ig*, *tf.*χ^2 and *tf.gr* is 86%, 79% and 81% macro-averaged F1 on SVM classification. Dai et al. [11] proposed the improved Chi-Square feature selection to solve many redundant feature problems. The macro-averaged F1 results of δ-χ^2 feature selection were 86.1% on SVM, 87.2% on Decision Tree (C4.5), 90.1% on Rough Set (RS). In our work, we propose to improve the supervised term weighting by applying Chi-Square feature selection [11]. The main idea of our new method is a term weighting by based-threshold δ-χ^2 term selection.

In terms of emotion class, we choose six emotions [12]; anger, disgust, fear, joy, sadness, and surprise. The rest of this paper is organized as follows: Section II describes the Term Extraction and Term Weighting Methods. Section III describes Framework for text classification. Section IV describes the Experiment and Evaluation. Finally, Section V describes Conclusion and Future Work.

II. TERM EXTRACTION

A. Preprocessing

The focus of our work is to identify emotions from texts found in social network. Therefore, the first step is to preprocess the data by transforming the crawled strings of characters into formal language. Then, we segment the formal language texts into terms before classifying them in the next steps.

B. Traditional term weighting method

In traditional term weighting method, document is segmented into terms [1]. Each term is added into a set of terms shown in (1)

$$T = \{t_1, t_2, t_3, \dots, t_m\} \tag{1}$$

where T is set of term, t_j is j^{th} term and m is number of term value. All documents are represented in (2)

$$D = \{d_1, d_2, d_3, \dots, d_N\} \tag{2}$$

where D is set of document, d_i is i^{th} document and N is number of documents. They are transformed to text representation method in Vector Space Model (VSM) [1]. Each term of document is represented by term weighting in (3)

$$\vec{\mathrm{d}}_i = (w_{i1}, w_{i2}, w_{i3}, \dots, w_{ij}) \tag{3}$$

where $\vec{\mathrm{d}}_i$ is i^{th} document vector, w_{ij} is j^{th} term weight on i^{th} document. Their vector is presented in (4) [1].

$$\begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \dots \\ \\ d_i \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & \dots & w_{1j} \\ w_{21} & w_{22} & w_{23} & \dots & w_{2j} \\ w_{31} & w_{32} & w_{33} & \dots & w_{31} \\ \dots & \dots & \dots & & \dots \\ \\ w_{i1} & w_{i2} & w_{i1} & \dots & w_{ij} \end{bmatrix} \tag{4}$$

1) Unsupervised Term Weighting Method

All terms of documents are represented by term weighting in VSM. Unsupervised term weighting measures the importance of this term and denotes how much this term distributes in the document. Example of unsupervised term weighting is binary, term frequency (*tf*), inverse document frequency (*idf*) and *tf.idf*. *tf* is number of term in document. *idf* is inverse number of document that term appears at least once. Then *tf.idf* is combined by term frequency and inverse number of document *tf.idf* weighting is presented in (5)

$$w_{tf.idf}(d_i, t_j) = tf(d_i, t_j) \times \log\left(\frac{N}{n(t_j)}\right) \tag{5}$$

This equation is *tf.idf* weighting method where $n(t_j)$ is number of document that j^{th} term appears in at least once. N is total number of documents in the training data.

2) Supervised Term Weighting Method

Supervised term weighting method analyzes the training data and produces co-occurrence frequency between term and emotion. Most of supervised term weighting are applied from term selection methods [10]. In traditional supervised term weighting method, *tf.rf* estimates occurrence of positive emotion of text compare with occurrence of negative emotion of text. The *tf.ig* measures the entropy variation between the presence and the absence of terms in a document while the *tf.*χ^2 measures the lack of independence between term and emotion. If *tf.*χ^2 value is zero, term and emotion are independence.

$$W_{Supervised}(t_j, d_i, e_k) = tf(t_j, d_i) \times W_{Supervised}(t_j, e_k) \tag{6}$$

Equation (6) is a general equation of supervised term weighting. Then, $W_{Supervised}(t_j, e_k)$ is replaced with equation (7)-(9) and (12) as follows:

$$W_{tf.rf} = \log\left(2 + \frac{P(t_j, e_k)}{P(t_j, \bar{e}_k)}\right) \quad (7)$$

$$W_{tf.ig} = \left[\begin{matrix} -P(e_k) \times \log P(e_k) + \\ P(t_j) \times P(e_k|t_j) \times \log P(e_k|t_j) + P(\bar{t}_j) \times P(e_k|\bar{t}_j) \times \log P(e_k|\bar{t}_j) \end{matrix}\right. \quad (8)$$

$$W_{\chi^2} = \frac{N \times \left[P(t_j, e_k) \times P(\bar{t}_j, \bar{e}_k) - P(t_j, \bar{e}_k) \times P(\bar{t}_j, e_k)\right]^2}{P(t_j) \times P(\bar{t}_j) \times P(e_k) \times P(\bar{e}_k)} \quad (9)$$

For equations (7)-(9), $P(t_j, e_k)$ is probability of documents that contains term t_j that belongs to emotion e_k. $P(t_j, \bar{e}_k)$ is probability of documents that contains term t_j that does not belong to emotion e_k. $P(\bar{t}_j, e_k)$ is probability of documents that does not contain term t_j that belongs to emotion e_k. $P(\bar{t}_j, \bar{e}_k)$ is probability of documents that does not contain term t_j that does not belong to emotion e_k. N is total number of document.

$$W_{Normalized}(d_i, t_j) = \frac{W(d_i, t_j)}{\sqrt{\sum_{j=1}^{m}\left[W^2(d_i, t_j)\right]}} \quad (10)$$

For equation (10), $W(d_i, t_j)$ is t_j term weight on document, d_i. $W_{Normalized}(d_i, t_j)$ is the t_j term normalized weight on documents, d_i. j is j^{th} order of term. m is number of term. Term weight normalization [1] is applied from cosine normalization to limit the term weight range between 0 and 1. This method can eliminate or prevent bias towards the document length effect.

C. Propose New Supervised Term Weighting Method

In this section, we propose new supervised term weighting method. This method places higher weight on terms that have matching emotions between owner and readers and that the terms are selected by human (readers.)

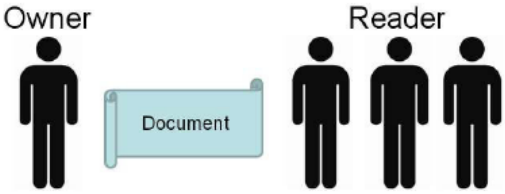


Figure 1. Supervised methods between owner and readers

In Equation (11), $P(T_j, E_k)$ is probability of human selected term T_j with matched emotion E_k. $P(T_j, \overline{E_k})$ is probability of human selected term T_j with no matced emotion, E_k. $P(\overline{T_j}, E_k)$ is probability of terms not selected by human T_j and matched emotion E_k. $P(\overline{T_j}, \overline{E_k})$ is probability of terms not selected by human T_j with no matched emotion E_k. N is total number of training documents. Then, we propose the based-threshold δ-χ^2 term selection for term weighting.

Figure 2. good classified term (left) and poor classified term (right)

Fig. 2 shows the nature of good and poor classified terms. Good classified term occurs when $P(T_j, E_k) \times P(\overline{T_j}, \overline{E_k}) > P(T_j, \overline{E_k}) \times P(\overline{T_j}, E_k)$ and poor classified term occurs when $P(T_j, E_k) \times P(\overline{T_j}, \overline{E_k}) \leq P(T_j, \overline{E_k}) \times P(\overline{T_j}, E_k)$. Then, we propose the based-threshold term selection for term weighting with equation (11) as follows:

$$\delta = \begin{cases} 1, & P(T_j, E_k) \times P(\overline{T_j}, \overline{E_k}) - P(T_j, \overline{E_k}) \times P(\overline{T_j}, E_k) > 0 \\ 0, & P(T_j, E_k) \times P(\overline{T_j}, \overline{E_k}) - P(T_j, \overline{E_k}) \times P(\overline{T_j}, E_k) \leq 0 \end{cases} \quad (11)$$

$$W_{tf.\delta-\chi 2} = tf \times \frac{\delta \times N \times \left[P(T_j, E_k) \times P(\overline{T_j}, \overline{E_k}) - P(T_j, \overline{E_k}) \times P(\overline{T_j}, E_k)\right]^2}{P(T_j) \times P(\overline{T_j}) \times P(E_k) \times P(\overline{E_k})} \quad (12)$$

For equation (11), Case δ = 1 if $P(T_j, E_k) \times P(\overline{T_j}, \overline{E_k}) - P(T_j, \overline{E_k}) \times P(\overline{T_j}, E_k) > 0$ and Case δ = 0 if $P(T_j, E_k) \times P(\overline{T_j}, \overline{E_k}) - P(T_j, \overline{E_k}) \times P(\overline{T_j}, E_k) \leq 0$. The δ value indicates whether term (T_j) is good or poor. For equation (12), δ value is combined with traditional χ^2 term weighting. Therefore, this equation is a term weighting by based-threshold δ-χ^2 term selection. Then, we present term weighting difference between tf.χ^2 and tf.δ-χ^2 in TABLE I-II as follows:

TABLE I. RESULT TERM WEIGHTING OF TF.CHI-SQUARE

Text	Word								
	ไม่	มาก	อะไร	มา	คน	มี	ว่า	ที่	...
1st Text	0	0.053	0	0	0.239528	0.960371	0	0.002524	...
2nd Text	2.801478	0	0	0	0	0.0184	0	0	...
3rd Text	24.06919	0	0	0	0	0	0.00564	3.432753	...
4th Text	0	0	17.30918	0	0	0.277257	2.653851	0.00756	...
5th Text	0	0	0	0.0242	1.09048	0	0	0	...
...	...	...	...	...	...	...	...	...	...

TABLE II. RESULT TERM WEIGHTING OF TF.DELTA-CHI-SQUARE

Text	Word								
	ไม่	มาก	อะไร	มา	คน	มี	ว่า	ที่	...
1st Text	0	0	0	0	0.239528	0.960371	0	0	...
2nd Text	0	0	0	0	0	0	0	0	...
3rd Text	0	0	0	0	0	0	0	3.432753	...
4th Text	0	0	17.30918	0	0	0.277257	2.653851	0	...
5th Text	0	0	0	0	1.09048	0	0	0	...
...	...	...	...	...	...	...	...	...	...

1st Text = ใช่|เธอ|เป็น|คน|ที่|บอก|รัก|ฉัน|ทำ|ให้|ฉัน|ดีใจ|และ|มี|ความ|สุข|มาก|

2nd Text = โกง|คอร์รัปชั่น||ประหาร|ชีวิต|คิด|ดุก|ตลอด|ชีวิต| |ไม่|มี|อภัยโทษ|เด็ดขาด|

3rd Text = ตอน|ที่|กล้อง|ส้ม|อ่ะ|ค่ะ| |ว่า|ตั้ง|นาน|ไม่|ส้ม| |อ่ะ| |เห็น|เงา| |ขาว|แวบ|

4th Text = มี|ใคร|พอ|ที่|จะ|ทราบ|มั้ย|ครับ|ว่า|มัน|คือ|อะไร|

5th Text = เอา|ไม้|ฟาด|หน้า|จอ|เมื่อ|เจอ|พวก|มือ|บอน|มา|โพสต์|ตอน|จบ|แกสัง|คน|

III. METHODOLOGY

The system is composed of two sections; the training, and the testing.

A. Data Training Section

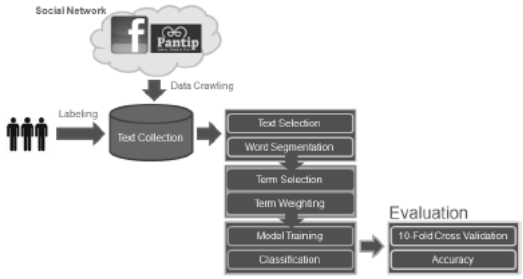


Figure 3. Framework for text classification

1) Data Crawling

Documents were crawled from social network such as Facebook and Pantip. Number of documents is 4,286 documents. Each document was obtained from one comment or post on social network.

2) Labeling

Persons were asked to label the closest emotion class for document. The choices are joy, fear, surprise, sadness, disgust, anger, neutral and others. Participants were separated into two groups. First group was the owners of the posts, comprising of 50 persons and second group was 50 persons who are the readers of the posts. Both groups were asked to select emotion labels. In addition, readers were asked to select emotional word from the documents. The user interface is shown in Fig. 4-5

Figure 4. User interface for emotion labeling

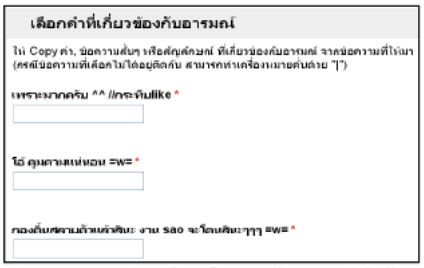


Figure 5. User interface for emotion term labeling

3) Text Selection

We selected documents which have the highest emotion voting scores. Documents in the training set must have the same three or four emotion votes from human while documents in the testing set must have the same four emotion votes from human.

4) Word Segmentation

Documents were segmented into words. We used TLex [13] Program using Conditional Random Fields (CRFs) algorithm. Then, each word was transformed to Bag of Word (BoW).

5) Term Selection

Since the data had high dimension of words, we performed word selection by word-emotion relevance scoring. We selected DF, IG, χ^2 and δ-χ^2 term selection for our experiment. These words are ranked by feature subset selection techniques

6) Term Weighting

The documents were transformed to BoW using Vector Space Model (VSM). Each word was assigned with term weights related to emotion. We chose Binary, *tf tf.idf*, *tf.rf*, *tf.ig*, *tf.*χ^2 and *tf.*δ-χ^2 term weighting mechanism.

7) Model Training

The final step of data training section was model training with the probability density function for a normal Gaussian distribution [14] ($g(x; \mu, \sigma)$). We present in equation (13) as follows:

$$g(x;\ \mu_e, \sigma_e) = \frac{1}{\sqrt{2\pi}\sigma_e} e^{-\frac{(x-\mu_e)^2}{2{\sigma_e}^2}} \qquad (13)$$

x is value of term t_j

μ_e is mean of x value

σ_e is standard deviation of x value

B. Data Testing Section

1) Classification

Classification model predicted the unlabeled emotion document from the training model. We utilized WEKA [15] to predict document with Naive Bayes algorithm. The equation can be written as follow:

$$e_{NB} = \underset{e_i \in E}{\operatorname{argmax}} \left\{ P(e_i) \prod_{j=1}^{m} P(t_j|e_i) \right\} \quad (14)$$

e_{NB} is predicted emotion with Naïve Bayes algorithm.
$P(e_i)$ is probability of document belong to i^{th} emotion.
$P(t_j|e_i)$ is probability of document contain j^{th} term given i^{th} emotion.
m is number of term value.
E is set of emotion class.
{ joy, fear, surprise, sadness, disgust, anger }

For equation (14), Naïve Bayes algorithm [14] is a simple probabilistic classifier based on applying Bayes's theorem with independence term assumptions.

2) Evaluation

System performance was measured using the K-fold cross validation and accuracy to test the dataset. Then, accuracy was a proportion number of document that is correctly classified and total number of test document in (15).

$$Accuracy(\%) = \frac{\#\ Document\ correctly\ classified}{\#\ Total\ Test\ Document} \times 100 \quad (15)$$

IV. EXPERIMENT AND EVALUATION

We performed experiment using a text collection from social network. As mentioned above, the total number of document was 4,286 documents from document's owner and reader of 100 persons. We started to collect documents from 8 October 2012 to 6 November 2012 using a PC Computer, CPU AMD Phenom(tm) II X6 1075T Processor 3.75GHz, RAM 4GB. Emotion class was joy, fear, surprise, sadness, disgust, anger. We used WEKA [15], an open-source machine learning program to text classification. Then, we selected Naïve Bayes classification [14] and used default setting. The result of k-fold cross validation and accuracy of the each ranked n^{th} term are presented in TABLE I-III and in FIGURE 6.

TABLE III. RESULT OF K-FOLD CROSS VALIDATION IN DIFFERENT TERM WEIGHTING

K-folds	Term Weighting	
	$tf{\cdot}\chi^2$	$tf{\cdot}\delta{\cdot}\chi^2$
4	63.69%	74.31%
7	63.27%	76.58%
10	64.33%	75.31%

TABLE III. summarizes K-fold cross validation from different term weighting methods. The results indicate that the performance of $tf{\cdot}\delta{\cdot}\chi^2$ term weighting method is superior than the $tf{\cdot}\chi^2$ term weighting method with the accuracy of 76.58%.

TABLE IV. RESULT ACCURACY OF DIFFERENT TERM WEIGHTING IN EACH RANKED N^{TH} TERM

Term Weighting	Number of n^{th} Ranked Term					
	10	40	160	640	2560	4260
Binary	27.44 %	35.10 %	46.90 %	48.97 %	50.15 %	51.03 %
tf	28.91 %	37.17 %	48.08 %	49.56 %	49.56 %	51.33 %
tf.idf	27.14 %	36.28 %	46.31 %	51.03 %	47.79 %	47.79 %
tf.rf	22.00 %	32.45 %	36.58 %	44.54 %	51.92 %	52.21 %
tf.ig	33.00 %	51.33 %	68.73 %	74.93 %	80.24 %	81.42 %
$tf{\cdot}\chi^2$	34.21 %	48.00 %	68.00 %	76.70 %	81.29 %	80.83 %
$tf{\cdot}\delta{\cdot}\chi^2$	34.21 %	52.00 %	67.00 %	74.93 %	88.13 %	88.79 %

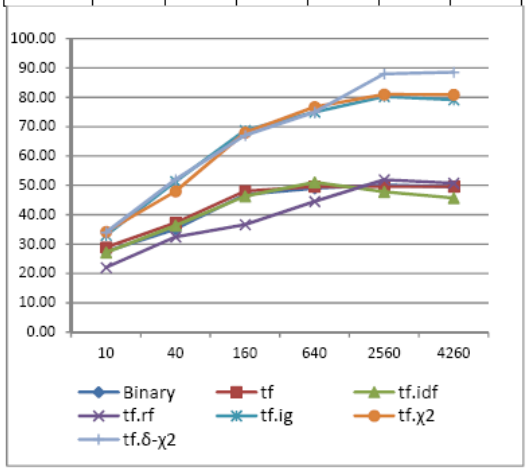


FIGURE 6. RESULT ACCURACY OF EACH RANKED N^{TH} TERM IN DIFFERENT TERM WEIGHTING

TABLE IV and Fig. 6, summarize accuracy of different term weighting methods in n^{th} rank word. The result indicates that accuracy of each term weighting is constant at 640^{th}-2560^{th} rank words. The best accuracy point on $tf{\cdot}\delta{\cdot}\chi^2$ weighting with Naïve Bayes classification is 88.79%.

TABLE V. RESULT ACCURACY OF DIFFERENT TERM WEIGHTING IN EACH TEXT COLLECTION

Text Collection	Term Weighting	
	$tf{\cdot}\chi^2$	$tf{\cdot}\delta\text{-}\chi^2$
Facebook	80.68%	84.09%
Pantip	79.60%	89.77%

TABLE V. summarizes the accuracy of different term weighting methods in each text collection. The $tf{\cdot}\delta\text{-}\chi^2$ term weighting method is superior to the $tf{\cdot}\chi^2$ term weighting method, approximately 3.41% on Facebook and 10.17% on Pantip text collection. The results reveal that the proposed method outperforms the traditional method on Facebook and Pantip text collection.

V. CONCLUSION AND FUTURE WORK

In this paper, we propose δ-Chi-Square supervised term weighting ($tf{\cdot}\delta\text{-}\chi^2$). Our method indicates that term weighting which selected by based-threshold $\delta\text{-}\chi^2$ term selection can improve classification performance. In addition, we propose ranking and feature subset selection techniques. The results show classification accuracy of 640^{th}-2560^{th} rank words to be equivalent to the classification accuracy of 4260 words.

Finally, the proposed δ-Chi-Square supervised term weighting ($tf{\cdot}\delta\text{-}\chi^2$) outperforms other term weighting methods including traditional Chi-Square supervised term weighting method. However our proposed method requires large training data set.

REFERENCES

[1] T. Danisman and A. Alpkocak, "Feeler: Emotion Classification of Text Using Vector Space Model", In AISB 2008 Convention, Communication, Interaction and Social Intelligence, Affective Language in Human and Machine, vol.2, 2008

[2] P. Inrak and S. Sinthupinyo, "Applying Latent Semantic Analysis To Classify Emotions In Thai Text", 2nd International Conference on Computer Engineering and Technology, 2010

[3] C. Lee and G. Geunbae Lee, "Emotion Recognition for Affective User Interfaces using Natural Language Dialogs", 16th IEEE International Conference on Robot & Human Interactive Communication, Jeju, Korea, August 26-29, 2007, 2007

[4] F. Liu, F. Liu, and Y. Liu, "A Supervised Framework for Keyword Extraction From Meeting Transcripts", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, MARCH 2011, pp. VOL. 19, NO. 3, 2011

[5] Y. Yang and J.O. Pedersen, "A Comparative Study on Feature Selection in Text Categorization", Proc. Int'l Conf. Machine Learning, pp. 412-420, 1997

[6] H. Wang, T. M. Khoshgoftaar and J. Van Hulse, "A Comparative Study of Threshold-based Feature Selection Techniques", IEEE International Conference on Granular Computing, 2010

[7] M. Lan, C. Lim Tan, Senior Member, IEEE, J. Su, and Y. Lu, "Supervised and Traditional Term Weighting Methods for Automatic Text Categorization", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, Apl. 2009

[8] N. Chirawichitchai, P. Sa-nguansat and P. Meesad, "Developing an Effective Thai Document Categorization Framework Based on Term Relevance Frequency Weighting", 2010 Eighth International Conference on ICT and Knowledge Engineering, 2010

[9] M. Lan, S. Yuan SUNG, H. Boon LOW and C. Lim TAN, "A Comparative Study on Term eighting Schemes for Text Categorization", Proceedings of International Joint Conference on Neural Networks, Montreal, Canada, July 31 - August 4, 2005

[10] F. Debole and F. Sebastiani, "Supervised Term Weighting for Automated Text Categorization," Proc. ACM Symp. Applied Computing, pp. 784-788, 2003.

[11] L. DAI, J. HU and W. Liu, "Using Modified CHI Square and Rough Set for Text Categorization with Many Redundant Features", 2008 International Symposium on Computational Intelligence and Design, 2008

[12] P. Ekman, "Facial Expressions of Emotion", American Psychological Association, Apl. 1993, pp. 376-379, 1992

[13] C. Haruechaiyasak and S. Kongyoung, "TLex: Thai Lexeme Analyser Based on the Conditional Random Fields", 2009

[14] Wray L. Buntine, "Operations for Learning with Graphical Models", Journal of Artificial Intelligence Research 2, Dec. 1994, pp. 159-225, 1994

[15] Weka Version 3.7.7 (c) 1999-2012, The University of Waikato, Hamilton, New Zealand, http://www.cs.waikato.ac.nz/~ml/weka/

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ประวัติผู้เขียนวิทยานิพนธ์

นายศุภกิจ เชื้อธนะภิญโญ เกิดวันอังคารที่ 3 พฤษภาคม พ.ศ. 2526 ที่จังหวัดสมุทรปราการ เข้าศึกษาระดับประถมศึกษา, มัธยมศึกษา และ สำเร็จการศึกษาระดับมัธยมศึกษาตอนปลายที่ โรงเรียนกรุงเทพคริสเตียนวิทยาลัย จังหวัดกรุงเทพมหานคร ในปี พ.ศ. 2544 ต่อมาได้เข้าศึกษาระดับปริญญาตรี และ สำเร็จการศึกษาระดับวิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้าสื่อสาร ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย จังหวัดกรุงเทพมหานคร ในปี พ.ศ. 2548 ต่อมาได้เข้าศึกษาระดับปริญญาโท ในหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย จังหวัดกรุงเทพมหานคร ในปี พ.ศ. 2553

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University