การวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทยโดยใช้เซ็นติกเน็ต

SENTIMENT ANALYSIS OF THAI CHILDREN STORIES USING SENTICNET

รัฐวุฒิ เฉิดสุขศักดา

RATHAWUT LERTSUKSAKDA

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2558

KMITL-2015-IT-M-001-014

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทยโดยใช้เซ็นติกเน็ต

SENTIMENT ANALYSIS OF THAI CHILDREN STORIES USING SENTICNET

รัฐวุฒิ เลิศสุขศักดา

RATHAWUT LERTSUKSAKDA

b.00266976
.1.....

สขทมู.....
สขทะเบียน 143965
วัน.เดือน.ปี 10 ต.ค. 2559

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2558

KMITL-2015-IT-M-001-014

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SENTIMENT ANALYSIS OF THAI CHILDREN STORIES USING SENTICNET

RATHAWUT LERTSUKSAKDA

A THESIS SUBMITTED IN FULFILLMENT OF
THE REQUIREMENT FOR THE DEGREE OF
MASTER OF SCHIENCE IN INFORMATION TECHONOLOGY
FACULTY OF INFORMATION TECHNOLOGY
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADGRABANG
2015

KMITL-2015-IT-M-001-014

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2015

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ใบรับรองวิทยานิพนธ์

หัวข้อวิทยานิพนธ์ การวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทยโดยใช้เซ็นติกเน็ต

Sentiment analysis of Thai children stories using SenticNet

นักศึกษา นายรัฐวุฒิ เลิศสุขศักดา

รหัสประจำตัว 57606145

ปริญญา วิทยาศาสตรมหาบัณฑิต

สาขาวิชา เทคโนโลยีสารสนเทศ

อาจารย์ที่ปรึกษาวิทยานิพนธ์ รองศาสตราจารย์ ดร.พรฤดี เนติโสภากุล

อาจารย์ที่ปรึกษาวิทยานิพนธ์ (ร่วม) ผู้ช่วยศาสตราจารย์ ดร.กิตติ์สุชาต พสุภา

คณะกรรมการสอบวิทยานิพนธ์	ลายมือชื่อ
รองศาสตราจารย์ ดร.วรพจน์ กรีสุระเดช	
ดร.เทพชัย ทรัพย์นิธิ	
รองศาสตราจารย์ ดร.พรฤดี เนติโสภากุล	
ผู้ช่วยศาสตราจารย์ ดร.กิตติ์สุชาต พสุภา	
ผู้ช่วยศาสตราจารย์ ดร.ธีรพงศ์ ลีลานุภาพ	
รองศาสตราจารย์ ดร.อาริต ธรรมโน	

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

วัน / เดือน/ ปี ที่สอบ วันอังคารที่ 15 ธันวาคม 2558 เวลา 13.00 น. เป็นต้นไป

สถานที่สอบ ณ ห้อง 333 ชั้น 3 คณะเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศรับรองแล้ว

(รองศาสตราจารย์ ดร.จันทรัปปรณ์ สถิตวิริยวงศ์)

คณบดีคณะเทคโนโลยีสารสนเทศ

วันที่ 4 **เดือน** ธันวาคม **พ.ศ.** 2558

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทยโดยใช้เซ็นติกเน็ต
นักศึกษา	นายรัฐวุฒิ เลิศสุขศักดา
รหัสนักศึกษา	57606145
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
พ.ศ.	2558
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร. พรฤดี เนติโสภากุล
	ผู้ช่วยศาสตราจารย์ ดร. กิตติ์สุชาต พสุภา

บทคัดย่อ

วิทยานิพนธ์นี้นำเสนอการวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทย โดยใช้เซ็นติกเน็ตสอง เป้าหมายคือ การแบ่งกลุ่มอารมณ์ของประโยคภาษาไทยที่อยู่ในนิทานเด็ก ออกเป็นเชิงบวก (Positive) เชิงลบ (Negative) หรือเป็นกลาง (Neutral) การวิจัยประกอบด้วย 3 ขั้นตอนใหญ่ ๆ ด้วยกัน ได้แก่ 1) การรวบรวม และจัดการสร้างชุดนิทาน 40 เรื่อง โดยแต่ละเรื่องจะมีการตัดประโยคโดยมนุษย์และแต่ประโยคจะมีการกำกับอารมณ์โดยมนุษย์ 2) การจัดสร้างทรัพยากรคำศัพท์ทางด้านอารมณ์ จากเซ็นติกเน็ตสอง โดยใช้เทคนิคการแปลแบบสองทาง 3) การออกแบบเฟรมเวิร์คเพื่อกำกับอารมณ์สำหรับแต่ละประโยคในนิทาน โดยใช้ซัพพอร์ตเวกเตอร์แมชชีน ผลลัพธ์จากการดำเนินการในการวิเคราะห์อารมณ์นั้น ได้ผลความแม่นยำดีที่สุดที่ 72.58 ± 1.08 เปอร์เซ็น การวิเคราะห์อารมณ์มีประโยชน์หลายประการ เช่น สามารถต่อยอดในงานทางด้านหุ่นยนต์ฮิวแมนนอยด์ หรือเป็นส่วนเสริมสำหรับระบบอ่านออกเสียงตามข้อความ (Text-to-Speech System) เพื่อให้การแสดงออกของระบบเหล่านั้น มีอารมณ์ความรู้สึกที่เป็นจริงมากยิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis	Sentiment Analysis of Thai children stories using SenticNet
Student	Mr. Rathawut Lertsuksakda
Student ID.	57606145
Degree	Master of Science
Program	Information Technology
Major	Information Science
Year	2558
The Advisors	Assc. Prof. Dr. Ponrudee Netisopakul
	Asst. Prof. Dr. Kitsuchart Pasupa

ABSTRACT

This thesis proposes a sentiment classification of Thai children stories using SenticNet-2. An objective is to develop a sentiment classification system that can classify a sentence's sentiment in Thai children stories into positive class, negative class, and neutral class. A research composes of three main tasks. 1) The task of collecting and building a dataset from 40 Thai children stories. Each story is manually sentence segmented and each sentence is manually sentiment tagged into 3 classes. 2) The task of building a Thai sentiment resource from senticnet-2 using a bidirectional translation technique. 3) The task of designing a framework for sentence-level sentiment classification using support vector machine. The best result of classification in this study is 72.58 ± 1.08%. There are many benefits of developing a sentiment analysis of Thai children stories, such as to develop a humanoid robot and to develop an add-on for a text-to-speech system; so that the expression of those systems can have more realistic feelings.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ ด้วยความกรุณาจาก รองศาสตราจารย์ ดร. พรฤดี เนติโสภากุล และ ผู้ช่วยศาสตราจารย์ ดร. กิติ์สุชาต พสุภา ซึ่งท่านทั้งสองได้สละเวลาอันมีค่าอย่างมากมาย เพื่อคอยสอน คอยให้ความรู้ คอยให้แง่คิด คอยให้คำเสนอแนะ มาด้วยดีโดยตลอด อีกทั้งยังช่วยตรวจสอบแก้ไขความเรียบร้อยของงานวิจัย ก่อนจะมีการเผยแพร่ออกสู่สาธารณชน

ขอขอบคุณคณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังที่ให้โอกาส โดยการสนับสนุนทุนการศึกษาเล่าเรียนในระดับบัณฑิตศึกษา และทุนสนับสนุนงานวิจัย อันเป็นสิ่งสำคัญอีกประการที่ขาดมิได้

สุดท้ายต้องขอขอบคุณบิดา มารดา และครอบครัว ที่ค่อยสนับสนุน และให้กำลังใจตลอดมา ขอขอบคุณเพื่อน ๆ พี่ ๆ และน้อง ๆ ในหน่วยปฏิบัติการการจัดการองค์ความรู้ และวิศวกรรมความรู้ (KMAKE Laboratory) ของคณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่คอยให้คำปรึกษา ให้กำลังใจ และช่วยเหลือกันในสิ่งที่เป็นความรู้ต่าง ๆ ตลอดมา

รัฐวุฒิ เลิศสุขศักดา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

บทคัดย่อ I

ABSTRACT II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VI

สารบัญรูปภาพ VII

บทที่ 1 บทนำ 8

1.1 ความเป็นมา และความสำคัญของปัญหา 8

1.2 ความมุ่งหมาย และวัตถุประสงค์ของการศึกษา 9

1.3 แนวคิดที่ใช้ในการวิจัย 9

1.4 การมีส่วนร่วมต่องานวิจัยในสาขาวิชา (Contribution) 9

1.5 ขอบเขตการวิจัย 10

1.6 ขั้นตอนการวิจัย 10

1.7 นิยามศัพท์ 10

1.8 ผลงานที่ได้รับการตีพิมพ์ 11

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 12

2.1 การประมวลผลภาษาธรรมชาติ (Natural Language Processing) 12

2.2 ทฤษฎีเกี่ยวกับอารมณ์ 13

2.3 การวิเคราะห์อารมณ์ (Sentiment Analysis) 18

2.4 งานวิจัยที่เกี่ยวข้อง 19

2.5 การเรียนรู้ของเครื่อง (Machine Learning) 26

2.6 ความแตกต่างระหว่างงานวิจัยนี้ และงานวิจัยที่เกี่ยวข้อง 29

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3 วิธีดำเนินการวิจัย 31
3.1 สมมติฐานการวิจัย 31
3.2 การรวบรวมชุดข้อมูลนิทาน 40 เรื่อง 32
3.3 การจัดสร้างทรัพยากรด้านอารมณ์ในภาษาไทย 35
3.4 การเตรียมข้อมูล (Preprocessing) 35
3.5 การเรียนรู้ของเครื่อง (Machine Learning) 42
บทที่ 4 ผลการทดลอง 46
4.1 ผลการรวบรวมทรัพยากรด้านอารมณ์ในภาษาไทย 46
4.2 ผลการทดลองจากการเรียนรู้ของเครื่อง (Machine Learning) 47
บทที่ 5 การวิเคราะห์ข้อผิดพลาด 58
5.1 ข้อผิดพลาดจากการกำกับอารมณ์ ให้กับประโยคในนิทาน 58
5.2 ข้อผิดพลาดที่เกิดขึ้นจากการสร้างฐานข้อมูลทรัพยากรด้านอารมณ์ 59
5.3 ข้อผิดพลาดที่เกิดขึ้นจากการดึงลักษณะสำคัญของประโยคที่ไม่มีค่าอารมณ์ 60
5.4 ข้อผิดพลาดที่เกิดขึ้นจากข้อจำกัดของเครื่องมือทางด้านการประมวลผลภาษาธรรมชาติ 60
บทที่ 6 สรุปผลการวิจัย และข้อเสนอแนะ 61
6.1 สรุปผลการวิจัย 61
6.2 ข้อเสนอแนะ 61
บรรณานุกรม 63
ภาคผนวก ก. ตัวอย่างทรัพยากรทางด้านอารมณ์ภาษาไทย 65
ภาคผนวก ข. ตัวอย่างนิทานเรื่องหนูบ้านกับหนูนา จากนิทาน 40 เรื่อง 77
ภาคผนวก ค. ผลงานวิจัยที่ได้รับการตีพิมพ์ 83
ประวัติผู้เขียน 95

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ 2.1 เปรียบเทียบระหว่างการวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทยโดยใช้เซ็นติกเน็ตและงานวิจัยที่เกี่ยวข้อง 29
ตารางที่ 3.1 ชนิดของคำที่ Jitar ได้ทำการแบ่งไว้ และความถี่ของคำแสดงอารมณ์ที่มนุษย์เลือกมา 37
ตารางที่ 3.2 คำต่าง ๆ ในประโยค พร้อมค่าอารมณ์เซ็นติก 38
ตารางที่ 3.3 คำต่าง ๆ ในประโยค พร้อมค่าอารมณ์เซ็นติก และประเภทของคำ 40
ตารางที่ 3.4 คำต่าง ๆ ในประโยคที่มีประเภทของคำคือ *vi*, *vt*, *adv*, *ncn* หรือ *adj* พร้อมค่าอารมณ์เซ็นติก และประเภทของคำ 41
ตารางที่ 3.5 คำที่มีค่า Polarity สูงสุด 6 อันดับแรกในประโยคที่มีประเภทของคำคือ *vi*, *vt*, adv, *ncn* หรือ *adj* พร้อมค่าอารมณ์เซ็นติก และประเภทของคำ 42
ตารางที่ 4.1 ตัวอย่างมโนทัศน์ "โทสะ" ซึ่งถูกแปลแบบสองทางมาจาก 7 มโนทัศน์ในภาษาอังกฤษ 46
ตารางที่ 4.2 สรุปผลการแปลเซ็นติกเน็ต 2 แบบสองทาง 47
ตารางที่ 4.3 แสดงค่าความแม่นยำของการทำนายผลโดยใช้การเรียนรู้ของเครื่อง 48
ตารางที่ 4.4 แสดงค่าความแม่นยำของการทำนายผลโดยใช้การเรียนรู้ของเครื่องแบบมี Auto Class Weight 49
ตารางที่ 4.5 แสดงค่าความแม่นยำของการทำนายผลโดยใช้การเรียนรู้ของเครื่องแบบมีการ Replacement 50
ตารางที่ 4.6 แสดงค่าความแม่นยำของการทำนายผลโดยใช้การเรียนรู้ของเครื่องแบบมี SMOTE.. 51
ตารางที่ 4.7 ผลลัพธ์ของสถิติทดสอบที 57

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ

ภาพที่ 2.1 วงล้ออารมณ์ของโรเบิร์ต พลุดชิค (Plutchik, 1987) 14
ภาพที่ 2.2 ระดับความเข้มข้นของอารมณ์ตามการแบ่งอารมณ์ของโรเบิร์ต พลุดชิค (Plutchik, 1987) 15
ภาพที่ 2.3 นาฬิกาทรายแห่งอารมณ์ (Cambria, Livingstone, & Hussain, The hourglass of emotions, 2012) 17
ภาพที่ 2.4 การผสมอารมณ์ระหว่างสองมิติของอารมณ์ (Cambria และ Hussain, Sentic Computing: Techniques, Tools, and Applications, 2012) 18
ภาพที่ 2.5 สถาปัตยกรรมระบบ pSenti (Mudinas, Zhang, และ Levene, 2012) 20
ภาพที่ 2.6 ผลลัพธ์จากระบบ *pSenti* ที่ถูกเรียกว่า Aspect-oriented Output 21
ภาพที่ 2.7 โครงสร้างของออนโทโลยีอารมณ์มนุษย์ (Cambria และ Hussain, Sentic Computing: Techniques, Tools, and Applications, 2012) 22
ภาพที่ 2.8 เฟรมเวิร์คของ S-Sense (Haruechaiyasak, Kongthon, Pornpimon, & Kanokorn, 2013) 24
ภาพที่ 2.9 POP เว็บแอพพลิเคชั่นวิเคราะห์อารมณ์ ขับเคลื่อนโดย S-Sense 26
ภาพที่ 2.10 ตัวอย่างของซัพพอร์ตสเปซที่มีการแบ่งข้อมูลออกจากกันโดยเส้นไฮเปอร์เพลน (Vapnik & Cortes, 1995) 27
ภาพที่ 3.1 กรอบแนวคิดการทำงานของการวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทยโดยใช้เซ็นติกเน็ต 32
ภาพที่ 3.2 ตัวอย่างการดึงลักษณะสำคัญ แบบเลือกสองคำเด่นที่แสดงอารมณ์ ของประโยคที่ว่า "การเคลื่อนไหวของต้นอ้อช่างสง่างามยิ่งนัก" (Lertsuksakda, Pasupa, และ Netisopakul, Sentiment Analysis on Thai Children Stories with Support Vector Machine, 2015) 45
ภาพที่ 4.1 เส้นโค้งอาร์โอซี สำหรับเคอร์เนล Linear 52
ภาพที่ 4.2 เส้นโค้งอาร์โอซี สำหรับเคอร์เนล Polynomial 53
ภาพที่ 4.3 เส้นโค้งอาร์โอซี สำหรับเคอร์เนล RBF 53
ภาพที่ 4.4 เมตริกซ์วัดประสิทธิภาพ สำหรับเคอร์เนล Linear 54
ภาพที่ 4.5 เมตริกซ์วัดประสิทธิภาพ สำหรับเคอร์เนล Polynomial 55
ภาพที่ 4.6 เมตริกซ์วัดประสิทธิภาพ สำหรับเคอร์เนล RBF 56

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมา และความสำคัญของปัญหา

การมีอยู่ของอินเทอร์เน็ต และผู้ใช้งานเป็นจำนวนมากในช่วงเวลาหลายปีที่ผ่านมา ก่อให้เกิดข้อมูลสารสนเทศจำนวนมหาศาลที่มีการไหลเวียนไปมาตลอดเวลา และมีปริมาณที่เพิ่มขึ้นเรื่อย ๆ อย่างไม่สามารถหยุดได้ ซึ่งส่วนหนึ่งของข้อมูลจำนวนมหาศาลเหล่านั้น ถูกจัดเก็บอยู่ในรูปแบบข้อความ (Text) ไม่ว่าจะอยู่ตามเว็บไซต์ เว็บบอร์ด บล็อก หรือสื่อสังคมออนไลน์ (Social Network) เช่น เฟสบุ๊ค (Facebook) หรือทวิตเตอร์ (Twitter) ก็ตามแต่ ซึ่งข้อมูลที่ถูกจัดเก็บอยู่ในรูปแบบข้อความนั้น เราสามารถจะกล่าวถึงมันได้ว่า ส่วนหนึ่งของมันคือสิ่งที่มีขั้วของอารมณ์แฝงอยู่ อันเป็นสิ่งเร้าที่เมื่อผู้อ่านได้อ่าน ย่อมต้องถูกชักจูงไปให้รู้สึกถึงอารมณ์ทางด้านใดด้านหนึ่ง และส่วนหนึ่งของมันอาจไม่ได้แฝงอารมณ์ใด ๆ ไว้ก็เป็นได้เช่นกัน ซึ่งอารมณ์ที่แฝงอยู่นั้น ก็คือสิ่งที่ผู้เขียนได้ถ่ายทอดออกมานั่นเอง

การรับรู้อารมณ์ที่แฝงอยู่ในข้อความนั้น ถือเป็นสิ่งที่ล้ำค่า และมีประโยชน์มาก ในหลาย ๆ ด้าน เช่น ในด้านของการบริหารความสัมพันธ์ลูกค้า (Customer relationship management) กรณีที่ผู้เป็นเจ้าของบริการ หรือสินค้า ต้องการทราบถึงผลตอบรับ (Feedback) ของผู้บริโภค เพื่อพัฒนาทางด้านการให้บริการ หรือพัฒนาคุณภาพของผลิตภัณฑ์ ว่าแง่มุมใดที่ขาดตกบกพร่องมีปัญหาควรปรับปรุง และแง่มุมใดที่ได้รับความพึงพอใจแล้ว หรือในด้านของการสร้างหุ่นยนต์ฮิวแมนนอยด์ (Humanoid robot) ที่สามารถรับรู้ สนทนา และตอบสนองต่ออารมณ์ของมนุษย์ได้ เป็นต้น

มันเป็นการง่ายที่เราจะรับรู้ถึงอารมณ์ที่แฝงอยู่ในข้อความใด ข้อความหนึ่ง โดยการที่เรานำมนุษย์มาอ่าน และบอกถึงสิ่งที่รับรู้ หรือรู้สึก จากการอ่านข้อความนั้น ๆ แต่คงไม่เป็นการดีนัก หากต้องตอบสนองกับข้อมูลจำนวนมาก ๆ แบบเวลาต่อเวลา นาทีต่อนาที วินาทีต่อวินาที ด้วยประการดังกล่าว การวิเคราะห์อารมณ์ (Sentiment analysis) จึงถูกพัฒนาขึ้นมา เพื่อตอบสนองต่อความต้องการ ในการระบุอารมณ์ของข้อความ สำหรับข้อมูลที่มีจำนวนมาก ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2 ความมุ่งหมาย และวัตถุประสงค์ของการศึกษา

1. เพื่อจัดสร้างฐานข้อมูลทรัพยากรทางด้านอารมณ์สำหรับภาษาไทย โดยใช้วิธีการแปลแบบสองทาง จากทรัพยากรทางด้านอารมณ์ภาษาอังกฤษที่ชื่อว่า SenticNet 2
2. เพื่อวิจัย และพัฒนาการวิเคราะห์อารมณ์ในนิทานภาษาไทย โดยใช้การเรียนรู้ของเครื่อง (Machine Learning) ที่สามารถทำนายอารมณ์ของประโยคในนิทาน โดยแบ่งอารมณ์เป็น 3 ประเภท คือ อารมณ์ขั้วบวก อารมณ์ขั้วลบ และอารมณ์ขั้วกลาง

1.3 แนวคิดที่ใช้ในการวิจัย

1. นาฬิกาทรายแห่งอารมณ์ (The Hourglass of Emotions)
2. การประมวลผลเซนติก (Sentic Computing)
3. เซนติกเน็ต (SenticNet)
4. การแปลแบบสองทาง (Bi-Directional Translation)
5. การเรียนรู้ของเครื่อง (Machine Learning)
6. ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

1.4 การมีส่วนร่วมต่องานวิจัยในสาขาวิชา (Contribution)

C1. งานวิจัยนี้ ได้จัดสร้างทรัพยากรทางด้านอารมณ์ในภาษาไทย โดยมอบให้กับทางห้องปฏิบัติการ KMAKE สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง สำหรับนำไปพัฒนาต่อยอด หรือนำไปใช้ในขอบเขตงานอื่น ๆ ต่อไป

C2. งานวิจัยนี้ ได้นำเสนอขั้นตอนการดึงลักษณะสำคัญ ของประโยคในนิทานภาษาไทย สำหรับการเรียนรู้ของเครื่องแบบมีผู้สอน

C3. งานวิจัยนี้ ได้ประเมิน และวิเคราะห์ข้อผิดพลาด ที่เกิดขึ้นในขั้นตอนต่าง ๆ อันได้แก่ ขั้นตอนการกำหนดขั้วของอารมณ์ให้กับนิทานแต่ละประโยชน์โดยใช้มนุษย์ ขั้นตอนการจัดสร้างทรัพยากรทางด้านอารมณ์ในภาษาไทย และขั้นตอนการดึงลักษณะสำคัญของนิทานแต่ละประโยค เพื่อเป็นประโยชน์สำหรับงานวิจัยอื่น ๆ ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.5 ขอบเขตการวิจัย

1. รวบรวมข้อมูลนิทานภาษาไทย 40 เรื่อง สำหรับงานวิจัยทางด้านการวิเคราะห์อารมณ์ โดยนำนิทานทั้ง 40 เรื่อง มาจาก (Netisopakul และ Wikaha, 2014) และทำการแบ่งประโยค รวมทั้งใช้มนุษย์ในการกำหนดค่าอารมณ์ให้กับแต่ละประโยค โดยมีการใช้มนุษย์เพียงคนเดียวเท่านั้นในการทำหน้าที่ดังกล่าว
2. จัดสร้างทรัพยากรทางด้านอารมณ์ในภาษาไทยด้วยวิธีการแปลแบบสองทาง (Vapnik และ Cortes, 1995) โดยใช้ SenticNet 2 (Cambria, Havasi, และ Hussain, SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis, 2012) เป็นข้อมูลตั้งต้น
3. แม้ว่าในข้อเท็จจริงแล้วนั้น อารมณ์ที่สามารถรับรู้ได้ จากการอ่านข้อความในนิทาน อาจจะมีเป็นจำนวนมาก แต่งานวิจัยนี้ จะทำการแบ่งประเภทของอารมณ์ ที่ได้จากการอ่านข้อความในนิทาน ออกเป็น 3 ประเภท คือ อารมณ์ขั้วบวก อารมณ์ขั้วลบ และอารมณ์ขั้วกลาง

1.6 ขั้นตอนการวิจัย

1. ศึกษา และค้นคว้า วิธีการวิเคราะห์อารมณ์ ที่มีอยู่ในปัจจุบัน
2. จัดสร้างฐานชุดข้อมูลนิทานภาษาไทย 40 เรื่อง สำหรับใช้ในงานวิจัยด้านการวิเคราะห์อารมณ์ในภาษาไทย
3. จัดสร้างทรัพยากรทางด้านอารมณ์ในภาษาไทยขึ้น สำหรับใช้ในงานวิจัยด้านการวิเคราะห์อารมณ์ในภาษาไทย
4. ทำการวิเคราะห์อารมณ์ โดยใช้แบบจำลองการเรียนรู้ของเครื่องแบบมีผู้สอน ทำการบันทึกและทำการวิเคราะห์ผลที่ได้รับ
5. วิเคราะห์ข้อผิดพลาด และนำเสนอทางแก้ไข ในขั้นตอนการวิจัยต่าง ๆ

1.7 นิยามศัพท์

1. มโนทัศน์ (Concept) หมายถึง ทรัพยากรทางด้านอารมณ์ 1 หน่วย จากฐานทรัพยากรทางด้านอารมณ์ ไม่ว่าจะเป็นคำศัพท์ (Word) หรือวลี (Phrase) ก็ตาม
2. เซ็นติก หมายถึง เวกเตอร์ของอารมณ์ที่ประกอบไปด้วยค่าต่าง ๆ ได้แก่ ค่าความพึงพอใจ (Pleasantness) ค่าความสนใจ (Attention) ค่าความอ่อนไหว (Sensitivity) ค่าการยอมรับ (Aptitude) ค่าขั้วของอารมณ์ (Polarity) ซึ่งแต่ละค่า จะถูกเก็บเป็นตัวเลขทศนิยม ที่อยู่ระหว่าง -1 ถึง 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.8 ผลงานที่ได้รับการตีพิมพ์

1. **Lertsuksakda, R.**, Netisopakul, P., Pasupa, K. (2014) Thai Sentiment Terms Construction using the Hourglass of Emotions. In: Proceeding of the 6th International Conference on Knowledge and Smart Technology (KST 2014), 30-31 January 2014, Chonburi, Thailand, pp 46-50.
2. **Lertsuksakda, R.**, Pasupa, K., Netisopakul, P., (2015) Sentiment Analysis on Thai Children Stories with Support Vector Machine In: Proceeding of the 20th International Symposium on Artificial Life and Robotics (AROB 2015), 21-23 Jan 2015, Beppu, Japan, pp 138-142

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะเป็นการกล่าวถึงทฤษฎี และงานวิจัยต่าง ๆ ที่เกี่ยวข้อง โดยมีจุดประสงค์เพื่ออธิบายถึงประเด็นต่าง ๆ ที่สำคัญต่อการศึกษา และการทดลองของงานวิจัยนี้

2.1 การประมวลผลภาษาธรรมชาติ (Natural Language Processing)

ภาษาธรรมชาติ ก็คือภาษาที่มนุษย์ใช้เพื่อการสื่อสาร ในชีวิตประจำวัน ไม่ว่าจะเป็นภาษาอังกฤษ ภาษาฮินดี ภาษาโปรตุเกส ภาษาอาหรับ ภาษาไทย หรือภาษาใด ๆ ก็ตาม ซึ่งภาษาธรรมชาติเหล่านี้นั้น มีความแตกต่างกับภาษาโปรแกรมมิ่ง หรือสัญลักษณ์ทางคณิตศาสตร์ เนื่องจาก ความหลากหลาย และการมีความเปลี่ยนแปลงอยู่เสมอ จากรุ่นหนึ่ง ยุคหนึ่ง สู่อีกรุ่นหนึ่ง อีกยุคหนึ่ง จึงยากที่จะกำหนดสิ่งที่เป็นกฎตายตัว ชัดเจน ให้กับภาษาธรรมชาติเหล่านั้น

โดยการประมวลผลภาษาธรรมชาติ เป็นศาสตร์หนึ่งที่เกี่ยวข้องกับด้านวิทยาศาสตร์คอมพิวเตอร์ ด้านการประมวลผลเกี่ยวกับภาษา และด้านปัญญาประดิษฐ์ ซึ่งกล่าวโดยสรุปได้ว่าเป็นศาสตร์ที่ครอบคลุมทุกประเภท ของการที่ทำให้คอมพิวเตอร์ สามารถจัดการกับภาษาธรรมชาติได้ (Bird, Ewan, & Loper, 2009)

2.1.1 การตัดคำ (Word Segmentation)

การตัดคำ เป็นขั้นตอนหนึ่งที่อยู่ในศาสตร์ด้านการประมวลผลภาษาธรรมชาติ ซึ่งเป็นการแยกคำแต่ละหน่วยในประโยคออกจากกัน เนื่องเพราะภาษาไทยนั้น ไม่มีเครื่องหมายบ่งบอกถึงขอบเขตของคำ (Word Boundary) ดังเช่นที่ภาษาอังกฤษมีช่องว่าง (Space) เป็นตัวบ่งบอกขอบเขตคำ การตัดคำ จึงเป็นสิ่งท้าทาย มีความสำคัญ และจำเป็นสำหรับภาษาที่ไม่มีเครื่องหมายบ่งบอกถึงขอบเขตอย่างเช่น ภาษาไทยนี้เอง เพื่อที่จะแยกคำออกมาจากประโยค และทำการประมวลผลทางภาษาธรรมชาติในขั้นตอนต่อ ๆ ไป

สำหรับเครื่องมือที่ใช้ในการตัดคำในภาษาไทยนั้น มีอยู่หลายเครื่องมือด้วยกัน เช่น SWATH (Smart Word Analysis for THai) จาก (Charoenpornsawat, 2003) หรือ KUCut จาก (Satayamas, 2009) ซึ่งตัว KUCut นั้น จะเป็นชุดคำสั่งในการตัดคำภาษาไทย สำหรับภาษา Python ที่ใช้เทคนิคการเรียนรู้ของเครื่องแบบไม่มีผู้สอน (Unsupervised Machine Learning) มาช่วยในการตัดคำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.2 การระบุชนิดของคำ (Part-of-speech Tagging)

การระบุชนิดของคำ เป็นขั้นตอนหนึ่งที่อยู่ในศาสตร์ด้านการประมวลผลภาษาธรรมชาติ เป็นกระบวนการที่ทำให้ระบบนั้น สามารถรับรู้ชนิด หรือหน้าที่ของคำศัพท์หนึ่ง ๆ ได้ เช่นว่า คำ ๆ นั้น เป็นคำนาม เป็นคำกริยา หรือคำประเภทอื่น ๆ

สำหรับตัวอย่างเครื่องมือที่ใช้ในการระบุชนิดของคำในภาษาไทยนั้น เช่น Jitar ที่ถูกเขียนมาด้วยภาษาจาวา (Java Programming Language) โดย ห้องปฏิบัติการวิจัยและประมวลผลภาษาธรรมชาติ (NAiST Laboratory) แห่งมหาวิทยาลัยเกษตรศาสตร์ (Jitar model and Jitar, 2010) โดย Jitar นั้น จะสามารถระบุกลุ่มชนิดของคำได้จำนวน 17 กลุ่ม และระบุชนิดของคำจำนวน 48 ชนิดด้วยกัน ดังแสดงไว้ในตารางที่ 3.1

2.1.3 การแปลแบบสองทาง (Bidirectional Translation)

การแปลแบบสองทาง (Wang และ Oard, 2006) มีเป้าหมาย เพื่อให้ได้รับคำแปลจากการแปลโดยใช้พจนานุกรม ที่ถูกต้องแม่นยำมากขึ้น และลดจำนวนคำแปลที่มีความหมาย (Sense) ไม่ตรงกับคำตั้งต้น ซึ่งวิธีการในการแปลแบบสองทางนั้น ก็คือ การนำคำตั้งต้นจากภาษาแรก แปลความหมายไปสู่ภาษาที่สอง และนำคำแปลที่เป็นภาษาที่สองแต่ละคำแปลนั้น แปลกลับไปสู่ภาษาแรก จากนั้นจึงทำการตรวจสอบว่า คำแปลภาษาที่สองคำใดบ้าง ที่สามารถแปลกลับไปสู่ภาษาแรก และมีคำแปลคำใดคำหนึ่ง ตรงกับคำตั้งต้นของภาษาแรกสุด จากนั้นจึงเลือกคำที่มีลักษณะเช่นนั้นออกมาเป็นคำแปลที่ต้องการ

2.2 ทฤษฎีเกี่ยวกับอารมณ์

2.2.1 นิยาม และการศึกษาเกี่ยวกับอารมณ์มนุษย์

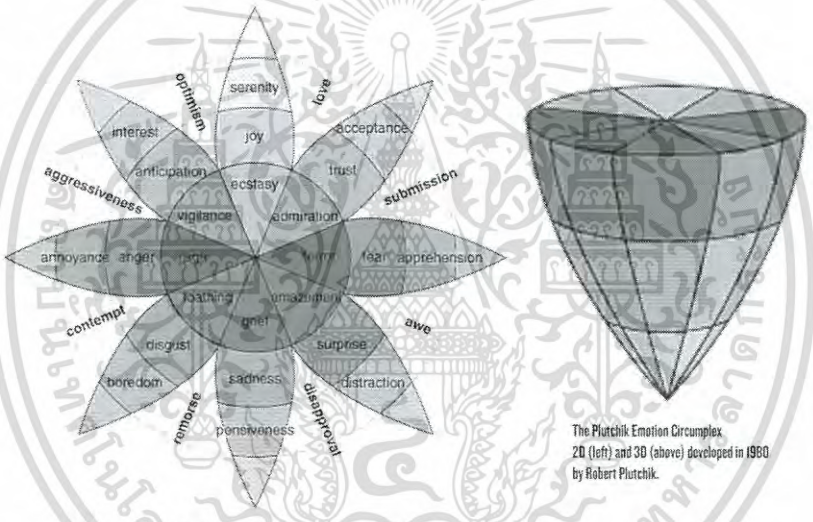
อารมณ์ คือ ภาวะที่เกิดจากการที่สิ่งเร้าภายนอกมากระตุ้น อันก่อเกิดซึ่งการตอบสนอง การแสดงออกมาในรูปแบบต่าง ๆ เช่น อาการชีพจรเต้นเร็ว อาการตื่นเต้น อาการหายใจรุนแรงขึ้น อาการหน้าแดงก่ำ เป็นต้น ซึ่งอารมณ์นั้น สามารถแสดงออกมาได้ในทั้งทางลบที่ไม่เป็นที่น่าพึงพอใจ และทางบวกที่เป็นที่น่าพึงพอใจ โดยอารมณ์นั้น จะเกี่ยวพันกับตำแหน่งในสมองบริเวณไฮโปธาลามัส (Hypothalamus) ที่จะมีจุดบนสมองเพื่อใช้ในการตอบสนองต่ออารมณ์ที่ต่างชนิดกัน ด้วยกันแปลข้อมูลจากสิ่งเร้าส่งเป็นกระแสประสาทไปยังประสาทมอร์เตอร์ (Motor Nerve) จนก่อเกิดอากัปกิริยาต่าง ๆ (Cambria, Livingstone, & Hussain, The hourglass of emotions, 2012)

การศึกษาเกี่ยวกับปรัชญาทางด้านอารมณ์นั้น เกิดขึ้นตั้งแต่สมัยกรีกโบราณ และโรมัน นักปรัชญาโรมัน ก่อนยุคคริสตกาล นามว่า มาร์คุส ทุลลิอุส คิเคโร หรือ มาร์เคิส ทัลลิเอิส ซิซีโร (Marcus Tullius Cicero) ได้นับจำนวน และแบ่งแยกอารมณ์ออกเป็น 4 ประเภทด้วยกัน คือ ความหวาดกลัว (metus) ความเจ็บปวด (Aegritudo) ความปรารถนา (Libido) และความยินดี (Laetitia)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

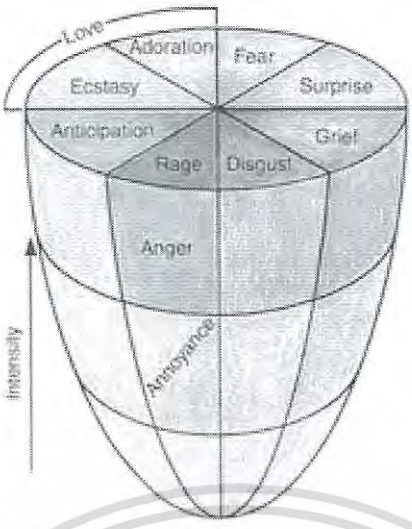
และภายหลังจากนั้น การศึกษาทฤษฎีวิวัฒนาการทางด้านอารมณ์ก็ได้เกิดขึ้นในช่วงท้ายของคริสต์ศตวรรษที่ 19 โดย ชาร์ล ดาวิน (Charles Darwin)

จาค แพงค์เซปป์ (Jaak Panksepp) ได้ทำการศึกษา และวิจัยในประเด็นที่เกี่ยวข้องกับอารมณ์ ซึ่งได้เสนอแนวคิดในการจำแนกประเภทต่าง ๆ ของอารมณ์ แบ่งออกเป็น 4 ประเภทด้วยกัน ได้แก่ ความคาดหวัง (Expectancy) ความเดือดดาล (Rage) การตื่นตระหนก (Panic) และความหวาดกลัว (Fear) ในขณะที่โรเบิร์ต พลุทชิค ได้ทำการแบ่งอารมณ์ออกเป็น 8 ชนิดด้วยกัน คือ ความกลัว (Fear) ความประหลาดใจ (Surprise) ความโศกเศร้า (Sadness) ความรังเกียจ (Disgust) ความโกรธ (Rage) ความคาดหวัง (Anticipation) ความรื่นเริง (Joy) และการยอมรับ (Acceptance) ซึ่งการจะแบ่งอารมณ์นั้น จึงเป็นเรื่องยาก และสามารถมีได้หลากหลายมุมมอง แล้วแต่ผู้ทำการแบ่งจะใช้จุดใดมาพิจารณาเป็นจุดในการอ้างอิง (ชูเดช, 2008)



ภาพที่ 2.1 วงล้ออารมณ์ของโรเบิร์ต พลุดชิค **(Plutchik, 1987)**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 2.2 ระดับความเข้มข้นของอารมณ์ตามการแบ่งอารมณ์ของโรเบิร์ต พลุดชิค **(Plutchik, 1987)**

2.2.2 นาฬิกาทรายแห่งอารมณ์ (The Hourglass of Emotions)

นาฬิกาทรายแห่งอารมณ์ (Cambria, Livingstone, & Hussain, The hourglass of emotions, 2012) คือแบบจำลองการจัดประเภทของอารมณ์ ที่ได้รับแรงบันดาลใจมาจากการศึกษาของพลุดชิค ในเรื่องอารมณ์มนุษย์ (Plutchik, 1987) ซึ่งเป็นการนำเอาแบบจำลองของพลุดชิคนั้น มาทำการอธิบายใหม่ ด้วยกับการแบ่งอารมณ์เป็นสี่ด้านที่ไม่ขึ้นต่อกัน แต่สามารถเกิดขึ้นพร้อม ๆ กันได้ โดยการอธิบายเช่นนี้ ได้รับแรงบันดาลใจมาจาก ทฤษฎีทางด้านสมองของมาร์วิน มินสกี (Minsky's Theory of the Mind) นักประชานศาสตร์ชาวอเมริกา นักวิทยาศาสตร์ ที่ทำการศึกษาเกี่ยวกับศาสตร์ด้านหนึ่ง ที่อยู่ในขอบข่ายของวิชาปัญญาประดิษฐ์ เขาเป็นหนึ่งในผู้ร่วมก่อตั้งห้องปฏิบัติการปัญญาประดิษฐ์ที่สถาบันเอ็มไอที (Massachusetts Institute of Technolog: MIT) โดยมาร์วินนั้นเชื่อว่าปฏิกิริยาในสมองของมนุษย์เรา เกิดมาจากการเปลี่ยนผัน เปลี่ยนแปลง ของทรัพยากรต่าง ๆ ที่อยู่ภายในสมอง โดยมีการเปิดชุดทรัพยากรหนึ่งจากเหล่าทรัพยากรเหล่านั้น ในขณะเดียวกันก็มีการปิดชุดทรัพยากรที่เหลือเอาไว้ อันก่อให้เกิดปฏิกิริยาในสมองรูปแบบต่าง ๆ ที่แตกต่างกัน ซึ่งอารมณ์มนุษย์นั้น ก็ถือว่าเป็นปฏิกิริยาหนึ่งที่เกิดขึ้นในสมองของมนุษย์ที่เกิดจากการเปิดปิด ชุดทรัพยากรเหล่านั้นนั่นเอง

ตามแบบจำลองนาฬิกาทรายแห่งอารมณ์นั้น จะมีการแบ่งอารมณ์ออกเป็น 4 ด้านด้วยกัน อันได้แก่ ด้านความพึงพอใจ (Pleasantness) ด้านความสนใจ (Attention) ด้านความอ่อนไหว (Sensitivity) และด้านการยอมรับ (Aptitude) โดยแต่ละด้าน หรือแต่ละมิตินั้น จะมีอยู่ด้วยกัน 6 ระดับ ซึ่งเรียกว่า ระดับเซ็นติก (Sentic Levels) อันเป็นสิ่งที่กำหนดความรุนแรง (Strength) ของการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

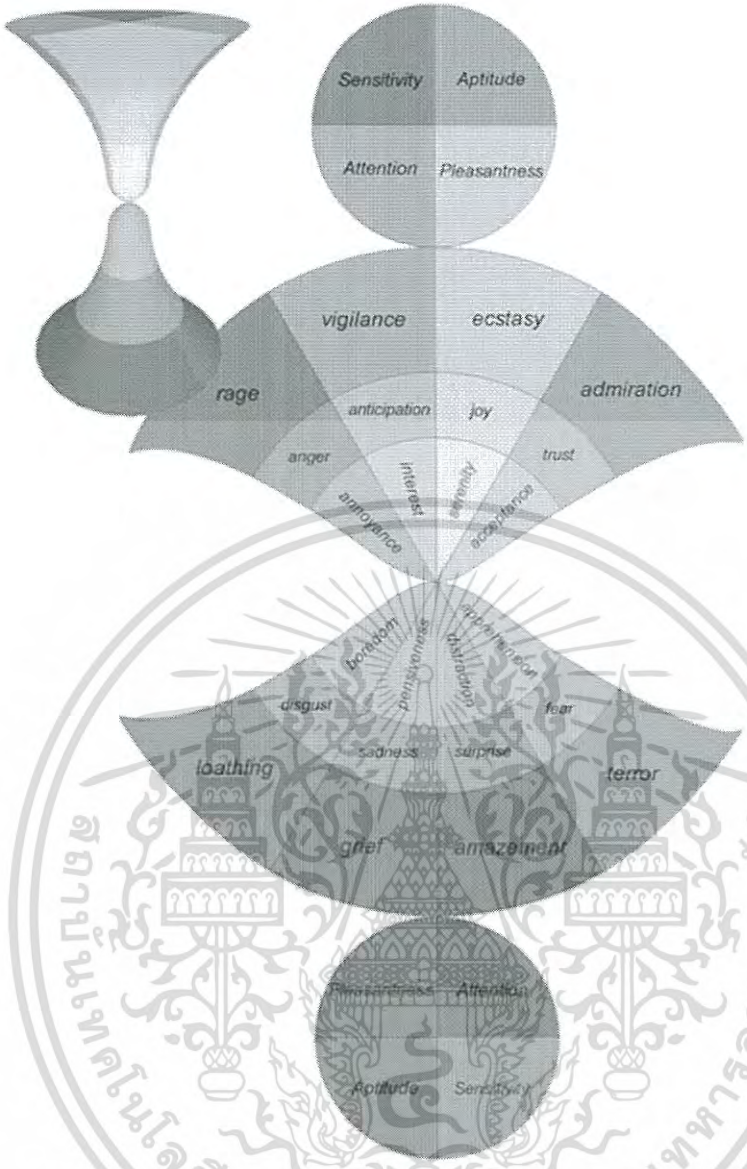
แสดงอารมณ์ในด้านนั้น ๆ โดยระดับของความรุนแรงนั้น จะอยู่ในขอบเขตตั้งแต่ค่า -3 จนถึงค่า +3 ซึ่งอารมณ์แต่ละด้านนั้น เป็นอิสระต่อกัน ทำให้สามารถเกิดขึ้นพร้อม ๆ กันได้ ดังแสดงการเกิดขึ้นพร้อมกัน ระหว่างอารมณ์ 2 ด้าน ได้ดังภาพที่ 2.4 ซึ่งการแทนค่าอารมณ์โดยอิงจากนาฬิกาทรายแห่งอารมณ์นั้น จะใช้เซ็นติกเวกเตอร์ ในการแทนค่าอารมณ์ทั้ง 4 ด้าน ในรูปแบบ [Pleasantness, Attention, Sensitivity, Aptitude]

เราสามารถยกตัวอย่างการแทนค่าอารมณ์ของความเดือดดาล (Rage) โดยใช้เซ็นติกเวกเตอร์ ได้ว่า [0, 0, +3, 0] ซึ่งก็หมายถึง มีค่า Sensitivity เป็น +3 ในขณะที่ค่า Pleasantness, Attention และ Aptitude เป็น 0 หรือสามารถยกตัวอย่างการแทนค่าอารมณ์ของความเศร้า (Sadness) ได้ว่า [-2, 0, 0, 0] ซึ่งก็หมายถึงมีค่า Pleasantness เป็น -2 ในขณะที่ค่า Attention, Sensitivity และ Aptitude เป็น 0 และสามารถยกตัวอย่างการแทนค่าอารมณ์ของความรัก (Love) ได้ว่า [+3, 0, 0, +3] ซึ่งก็หมายถึงมีค่า Pleasantness และ Aptitude เป็น +3 ในขณะที่มีค่า Attention และ Sensitivity เป็น 0 เป็นต้น

สำหรับตัวอย่างของอารมณ์อันเกิดมาจากค่าเพียงมิติเดียว เช่นคำว่า ความเดือดดาล (Rage) และความเศร้า (Sadness) นั้น จะแสดงให้เห็นในภาพที่ 2.3 ในขณะที่ตัวอย่างของอารมณ์อันเกิดมาจากค่าสองมิติ หรืออารมณ์ 2 ด้าน จะแสดงให้เห็นใน ภาพที่ 2.4

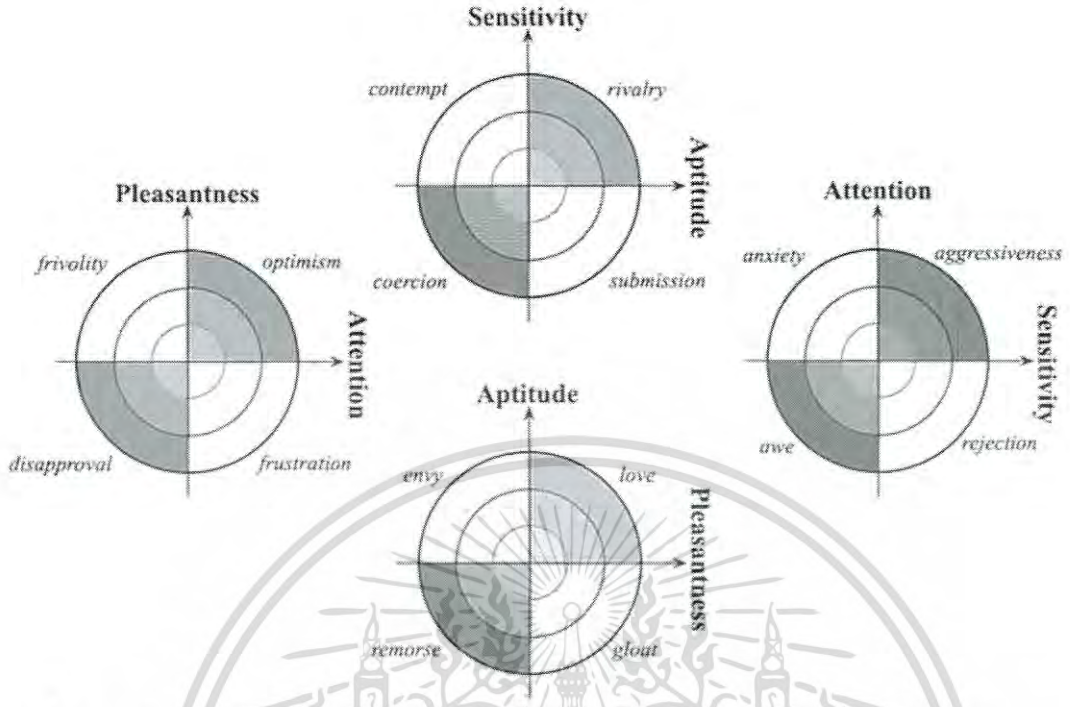
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง



ภาพที่ 2.3 นาฬิกาทรายแห่งอารมณ์ **(Cambria, Livingstone, & Hussain, The hourglass of emotions, 2012)**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
143965



ภาพที่ 2.4 การผสมอารมณ์ระหว่างสองมิติของอารมณ์ **(Cambria และ Hussain, Sentic Computing: Techniques, Tools, and Applications, 2012)**

2.3 การวิเคราะห์อารมณ์ (Sentiment Analysis)

การวิเคราะห์อารมณ์ คือศาสตร์หนึ่งทางด้านการประมวลผลภาษาธรรมชาติ ที่มีจุดมุ่งหมายในการดึงอารมณ์ความรู้สึกออกมาจากข้อความ (Text) เพื่อแสดงขั้วของอารมณ์ (Polarity) หรืออาจจะมีการแสดงระดับความเข้มของของอารมณ์ (Strength) ด้วย ก็เป็นได้ ซึ่งการศึกษาศาสตร์ด้านนี้ในปัจจุบัน จะมีความเกี่ยวข้องกันอย่างมากกับการเรียนรู้ของเครื่อง (Machine Learning) สำหรับการวิเคราะห์อารมณ์นั้น ยังมีชื่อเรียกอื่น ๆ อีกมากมายหลายชื่อ ไม่ว่าจะเป็น การทำเหมืองความคิดเห็น (Opinion Mining) ความรู้สึกนึกคิด (Subjectivity) การวิเคราะห์จุดยืน (Analysis of Stance) การประเมินค่า (Appraisal) จุดประสงค์ของความคิดเห็น (Point of View) สิ่งอันเป็นเครื่องแสดง (Evidentiality) หรือชื่ออื่น ๆ นอกจากนี้ (Mudinas, Zhang, & Levene, 2012)

งานพื้นฐานของการวิเคราะห์อารมณ์ คือการจำแนกขั้วของอารมณ์ ให้กับข้อความที่ได้รับเข้ามา ซึ่งอาจจะมีการวิเคราะห์ในระดับเอกสาร (Document Level) ระดับประโยค (Sentence Level) หรือระดับแง่มุม มุมมอง (Aspect Level) ว่าขั้วอารมณ์นั้น เป็นไปในทางบวก (Positive Polarity) ทางลบ (Negative Polarity) หรือค่ากลาง (Neutral Polarity) ซึ่งในระดับขั้นสูง ขั้นก้าวหน้า ของงานทางด้านการวิเคราะห์อารมณ์นั้น อาจจะมีการเพิ่มสถานะทางอารมณ์เข้ามากำกับ เช่น หากอารมณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ของข้อความเป้าหมายเป็นไปในทางลบ ก็จะมีการเพิ่มสถานะทางอารมณ์ ให้รายละเอียดลงไปอีกว่าเป็นอารมณ์ "เศร้า (Sad)" หรืออารมณ์ "โกรธ (Angry)" หรือหากอารมณ์ขอข้อความเป้าหมายเป็นไปในทางบวก ก็จะมีการเพิ่มสถานะทางอารมณ์ให้รายละเอียดลงไปอีกเช่นกันว่า เป็นอารมณ์ "พึงพอใจ (Pleasantness)" หรืออารมณ์ "มีความสุข (Happy)" เป็นต้น

2.3.1 การประมวลผลแบบเซ็นติก (Sentic Computing)

การประมวลผลแบบเซ็นติกนั้น (Cambria & Hussain, Sentic Computing: Techniques, Tools, and Applications, 2012) เป็นแนวคิดแบบหลากหลายสาขาวิชา (Multi-disciplinary) สำหรับการวิเคราะห์อารมณ์ เพื่อให้มีประสิทธิภาพสูงสุด ในการตรวจจับ อธิบาย และประมวลผลความคิดเห็นที่อยู่บนโลกอินเทอร์เน็ต โดยวิชาที่ประกอบขึ้นมาเป็นการประมวลผลแบบเซ็นติกนั้น คือ ทั้งศาสตร์ทางด้านคอมพิวเตอร์ และศาสตร์ทางด้านสังคม

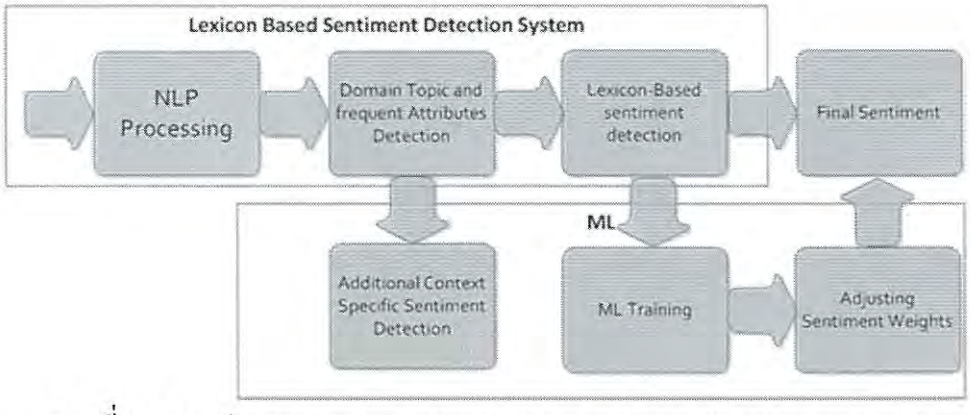
การประมวลผลเซ็นติกนั้น จะเกี่ยวข้องกับศาสตร์ต่าง ๆ มากมาย เช่น มีการใช้เทคนิคในเรื่องปัญญาประดิษฐ์ (Artificial Intelligence) และเว็บเชิงความหมาย (Semantic Web) ในแทนความรู้และการอนุมาน มีการใช้วิชาคณิตศาสตร์ (Mathemathics) เพื่อดำเนินการในงานส่วนต่าง ๆ เช่น การทำเหมืองกราฟ (Graph Mining) การลดมิติของข้อมูลที่มีหลายมิติ (Multi-dimensionality Reduction) มีการใช้วิชาภาษาศาสตร์ (Linguistics) เพื่อทำการวิเคราะห์ภาษาระดับข้อความ (Discourse Analysis) และวัจนปฏิบัติศาสตร์ (Pragmatics) มีการใช้วิชาจิตวิทยา (Psychology) เพื่อทำความเข้าใจเกี่ยวกับการรับรู้ของมนุษย์ และเจตคติ มีการใช้วิชาสังคมวิทยา (Sociology) เพื่อทำความเข้าใจถึงความผันแปรของสื่อสังคม และสิ่งที่โน้มน้าวสังคม และมีการใช้วิชาจริยธรรม (Ethics) เพื่อทำความเข้าใจประเด็นที่เกี่ยวข้องกับธรรมชาติของความคิดของมนุษย์ และใช้ในการสร้างเครื่องที่มีอารมณ์ความรู้สึก (Emotional Machine) เป็นต้น

2.4 งานวิจัยที่เกี่ยวข้อง

2.4.1 ระบบวิเคราะห์อารมณ์ *pSenti*

ระบบ *pSenti* เป็นระบบวิเคราะห์อารมณ์ที่ถูกพัฒนาขึ้นโดย (Mudinas, Zhang, และ Levene, 2012) แห่งมหาวิทยาลัยลอนดอน (University of London) ผ่านบทความที่มีหัวข้อว่า "การผสมผสานกันระหว่างแนวคิดฐานพจนานุกรม กับแนวคิดฐานการใช้การเรียนรู้ของเครื่อง สำหรับการวิเคราะห์อารมณ์ในระดับมโนทัศน์ (Combining Lexicon and Learning Based Approaches for Concept-level Sentiment Analysis) โดยระบบนี้ จะรับข้อมูลเข้าในรูปแบบไฟล์ข้อความ (Text File) ของบทความที่ต้องการนำมาวิเคราะห์ ซึ่งผลลัพธ์จากการวิเคราะห์นั้นจะออกมาในรูปแบบแง่มุมที่ถูกกล่าวถึง (Aspect) และค่าอารมณ์ที่แง่มุมนั้นได้รับ หรือความคิดเห็นต่อแง่มุมนั้น (View)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 2.5 สถาปัตยกรรมระบบ pSenti **(Mudinas, Zhang, และ Levene, 2012)**

การทำงานของ *pSenti* นั้น ประกอบไปด้วยขั้นต่อต่าง ๆ 7 ขั้นตอนด้วยกัน ได้แก่

1. **NLP Processing** เป็นขั้นตอนทางการประมวลผลภาษาธรรมชาติ โดยจะมีการใช้ Stanford CoreNLP Toolkit ในการทำการกำกับประเภทของคำ (Part-of-speech Tagging) ให้กับข้อมูลรับเข้า
2. **Domain Topic and Frequent Attributes Detection** เป็นขั้นตอนการเลือกแง่มุมที่ถูกกล่าวถึง (Aspect) มาเป็นผู้ท้าชิง (Candidate) โดยแง่มุมที่ถูกกล่าวถึงนั้น จะพิจารณาจากคำที่เป็นคำนาม ซึ่งได้รับการกำกับประเภทของคำมาในขั้นตอนก่อนหน้า โดยระบบจะทำการตัดคำที่ไม่สำคัญออก (Stop Word) ตัดชื่อเฉพาะต่าง ๆ ออก และตัดคำศัพท์ทางด้านอารมณ์ที่รู้อยู่แล้วออกไป จากนั้นจะทำการตัดผู้ท้าชิง ที่มีความถี่ในการปรากฏออกมา น้อยกว่าจำนวน 5 ครั้ง หรือผู้ท้าชิงที่ถูกพบก่อนหน้าแล้ว ออกไป แล้วจึงใช้พจนานุกรมคำเหมือน (Lexical Similarity) จับกลุ่มผู้ท้าชิงที่มีความหมายเดียวกัน จากนั้นจึงสรุปรายการแง่มุมที่ถูกกล่าวถึง (Aspect) ที่เป็นผู้ท้าชิงใน 100 อันดับแรก และรวมเข้ากับ 10 อันดับสูงสุดของความเห็นของแต่ละแง่มุมที่ถูกเลือก
3. **Context Specific Sentiment Detection** เป็นขั้นตอนการแทนที่สำนวน (Idiom) หรือสัญรูปอารมณ์ (Emoticon) ที่บ่งบอกถึงอารมณ์ที่รู้กัน ด้วยกับคำจำลอง (Pseudo-word) เช่น _Good_One_ หรือ _Bad_Three_ เป็นต้น
4. **Lexicon-Based Sentiment Detection** เป็นขั้นตอนการกำกับอารมณ์ โดยใช้ทรัพยากรทางด้านอารมณ์ที่ได้เตรียมไว้
5. **ML Training** เป็นขั้นตอนการเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised Machine Learning) โดยจะมีการดึงลักษณะสำคัญออกมา (Feature extraction) และทำการฝึกสอนให้กับข้อมูลสำหรับการฝึกสอน จากนั้นจะมีการหาค่าสัมประสิทธิ์ของซัพพอร์ตเวกเตอร์แมชชีน (SVM Coefficient) เพื่อเก็บไว้ใช้ในขั้นตอนถัดไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6. **Adjusting Sentiment Weights** เป็นขั้นตอนการปรับเปลี่ยนน้ำหนักของอารมณ์ให้เหมาะสมขึ้น โดยใช้ค่าสัมประสิทธิ์ของซัพพอร์ตเวกเตอร์แมชชีน ที่ได้มาจากขั้นตอนการเรียนรู้ของเครื่อง
7. **Final Sentiment** เป็นการคิดค่าของอารมณ์ครั้งสุดท้าย เพื่อให้ค่าอารมณ์นั้น อยู่ในพิสัย (Range) ที่ต้องการคือ ระหว่าง -1 ถึง 1 โดยใช้สมการดังสมการที่ 2.1 หรือระหว่าง 1 ถึง 5 โดยใช้สมการดังสมการที่ 2.2 โดยกำหนดให้ pos และ neg คือค่าทั้งหมดของอารมณ์ด้านบวก และอารมณ์ด้านลบตามลำดับ

$$S_{senti} = \frac{1}{2}\log_2 \frac{pos}{neg} \tag{2.1}$$

$$S_{stars} = 2 \cdot S_{senti} + 3 \tag{2.2}$$

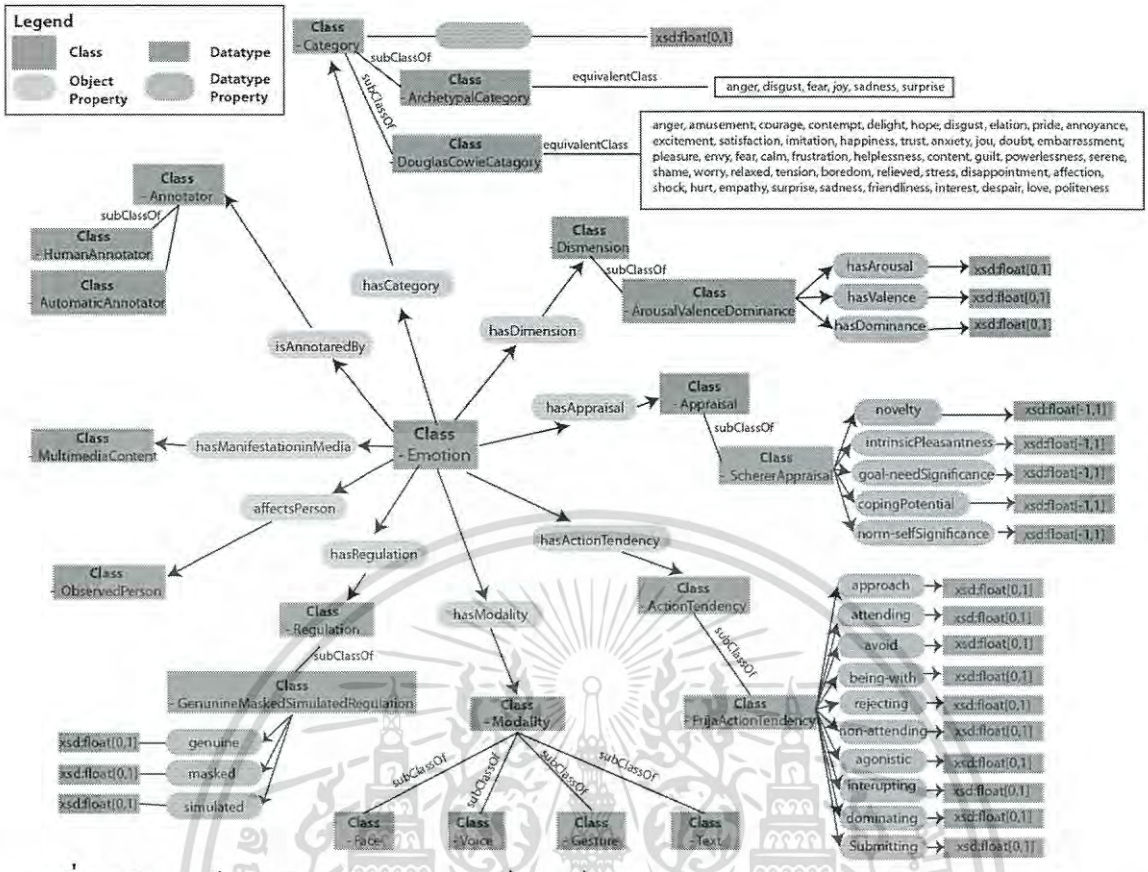
Customer Review => { (Aspect$_1$: View$_1$), (Aspect$_2$: View$_2$), . . . , (Aspect$_k$: View$_k$) },
e.g.,
A user comment on Google Chrome => { (Appearance: +0.8), (Plugins: +0.6), . . . , (Speed: +0.9) }.

ภาพที่ 2.6 ผลลัพธ์จากระบบ *pSenti* ที่ถูกเรียกว่า Aspect-oriented Output

2.4.2 ออนโทโลยีอารมณ์มนุษย์ (Human Emotions Ontology: HEO)

ออนโทโลยีอารมณ์มนุษย์ คือออนโทโลยี ที่เป็นฐานความรู้อันเกี่ยวข้องกับอารมณ์ของมนุษย์ ซึ่งถูกพัฒนามาจากภาษาโอดับเบิ้ลยูแอล (The W3C Web Ontology Language: OWL) ซึ่งวางรากฐานการแบ่งอารมณ์บนทฤษฎีอารมณ์ของพลุคชิค แสดงโครงสร้างได้ดังภาพที่ 2.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 2.7 โครงสร้างของออนโทโลยีอารมณ์มนุษย์ **(Cambria และ Hussain, Sentic Computing: Techniques, Tools, and Applications, 2012)**

2.4.3 เซ็นติกเน็ต 2 (SenticNet 2)

เซ็นติกเน็ต 2 (Cambria, Havasi, & Hussain, SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis, 2012) คือ ทรัพยากรทางด้านอารมณ์ที่เก็บเป็นภาษาอังกฤษ อันเป็นผลลัพธ์การสร้างทรัพยากร ที่ตั้งบนศาสตร์การประมวลผลเซ็นติก ซึ่งพัฒนาขึ้นมาจาก เซ็นติกเน็ตเวอร์ชั่นแรก โดยการกล่าวถึง เซ็นติกเน็ต 2 นั้น ก็คือการกล่าวถึงการวิเคราะห์อารมณ์ในระดับมโนทัศน์ (Concept-level Sentiment Analysis)

เซ็นติกเน็ต 2 นั้น คือทรัพยากรทางด้านอารมณ์ที่ถูกจัดเก็บอยู่ในรูปแบบอาร์ดีเอฟ (Resource Description Framework: RDF) อันเป็นมาตรฐานที่วางอยู่บนรูปแบบของเอกซ์เอ็มแอล (EXtensible Markup Language: XML) อีกทีหนึ่ง ที่ใช้เพื่อการเข้ารหัสความรู้ ซึ่งเป็นมาตรฐานของดับเบิลยูทรีซี (World Wide Web Consortium: W3C) ซึ่งการเก็บค่าอารมณ์สำหรับแต่ละมโนทัศน์ในเซนติกเน็ตนั้น จะอิงอยู่กับนาฬิกาทรายแห่งอารมณ์ โดยแต่ละมโนทัศน์ จะมีค่าอารมณ์ด้วยกัน 4 ด้าน อันได้แก่ ด้านความพึงพอใจ (Pleasantness) ด้านความสนใจ (Attention) ด้านความอ่อนไหว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Sensitivity) และด้านการยอมรับ (Aptitude) และมีค่าขั้วของอารมณ์ (Polarity) อีก 1 ค่า และมีการเก็บความสัมพันธ์ไปยังมโนทัศน์ที่มีความหมายใกล้เคียงกัน (Semantic Similarity) อีกด้วย

ตัวอย่างของมโนทัศน์ "Wrath" ใน SenticNet 2 ที่เขียนอยู่ในรูปแบบ RDF สามารถแสดงได้ดังนี้

```
<rdf:Description rdf:about="http://sentic.net/api/en/concept/wrath">
 <rdf:type rdf:resource="http://sentic.net/api/concept"/>
 <text xmlns="http://sentic.net/api/">wrath</text>
 <semantics xmlns="http://sentic.net/api/"
       rdf:resource="http://sentic.net/api/en/concept/temper"/>
 <semantics xmlns="http://sentic.net/api/"
       rdf:resource="http://sentic.net/api/en/concept/madness"/>
 <semantics xmlns="http://sentic.net/api/"
       rdf:resource="http://sentic.net/api/en/concept/scorn"/>
 <semantics xmlns="http://sentic.net/api/"
       rdf:resource="http://sentic.net/api/en/concept/despise"/>
 <semantics xmlns="http://sentic.net/api/"
       rdf:resource="http://sentic.net/api/en/concept/disgust"/>
 <pleasantness xmlns="http://sentic.net/api/"
       rdf:datatype="http://www.w3.org/2001/XMLSchema#float">-0.737</pleasantness>
 <attention xmlns="http://sentic.net/api/"
       rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.0</attention>
 <sensitivity xmlns="http://sentic.net/api/"
       rdf:datatype="http://www.w3.org/2001/XMLSchema#float">+0.899</sensitivity>
 <aptitude xmlns="http://sentic.net/api/"
       rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.0</aptitude>
 <polarity xmlns="http://sentic.net/api/"
       rdf:datatype="http://www.w3.org/2001/XMLSchema#float">-0.545</polarity>
</rdf:Description>
```

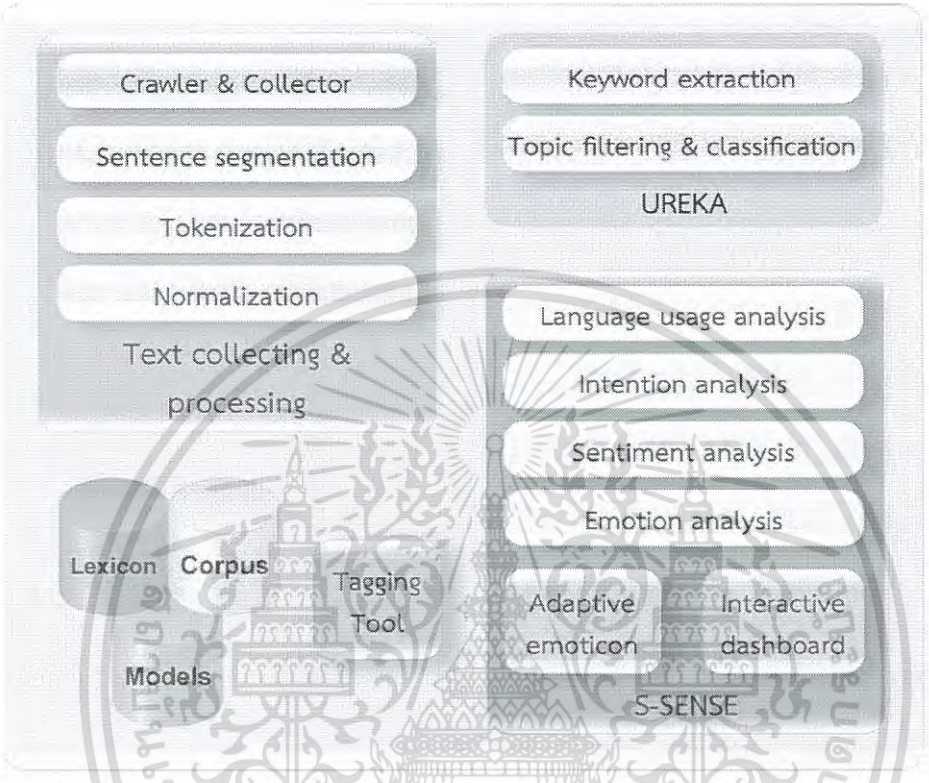
2.4.4 ระบบวิเคราะห์อารมณ์ S-Sense: Social Sensing

ระบบวิเคราะห์อารมรณ์ S-Sense เสนอโดย (Haruechaiyasak, Kongthon, Pornpimon, & Kanokorn, 2013) เป็นเฟรมเวิร์คการวิเคราะห์อารมณ์ (Sentiment Analysis Framework) สำหรับเครือข่ายสังคมออนไลน์ เพื่อทำการรวบรวมข้อความบนเว็บไซต์เครือข่ายสังคมออนไลน์ และเว็บบอร์ดต่าง ๆ ทำการติดตาม และทำการวิเคราะห์ ข้อมูลเหล่านั้น ซึ่งภาษาที่ใช้ตามสื่อสังคมออนไลน์หรือเว็บบอร์ดนั้น มักจะเป็นภาษาแบบไม่เป็นทางการ โดยที่ S-Sense นั้น ได้รับการออกแบบมาเพื่อรับมือกับภาษาที่ไม่เป็นทางการ หรือภาษาพูดเหล่านี้ ให้ได้อย่างมีประสิทธิภาพ

ปัจจุบัน ข้อมูลในรูปแบบข้อความจำนวนมาก ตามสื่อสังคมออนไลน์ หรือเว็บบอร์ด ที่กล่าวถึงผลิตภัณฑ์ กล่าวถึงสินค้า ไม่ว่าจะในทางด้านบวก ด้านแนะนำ ส่งเสริม หรือในทางด้านลบ ด้านตำหนิ ติเตียน ก็ตาม ซึ่ง S-Sense นั้น สามารถรับมือกับข้อมูลเหล่านี้ เพื่อทำการรวบรวม ติดตาม และวิเคราะห์ ทัศนคติของผู้ที่เป็นลูกค้า ว่ารู้สึกอย่างไรต่อผลิตภัณฑ์หนึ่ง ๆ ได้ เพื่อนำสิ่งที่ได้รับจากการวิเคราะห์ดังกล่าวนั้น มาพัฒนาคุณภาพของสินค้า หรือพัฒนาทางด้านการบริหารลูกค้าสัมพันธ์ เพื่อความได้เปรียบทางการตลาดต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระบบปัจจุบันของ S-Sense นั้น จะเป็นการนำอัลกอริทึมของนาอีฟเบย์แบบมัลติโนเมียล มาใช้ (Multinomial Naïve Bayes) เนื่องจากว่า เป็นอัลกอริทึมที่ใช้ข้อมูลสำหรับการฝึกสอนจำนวนน้อย เพื่อทำการประมาณค่าพารามิเตอร์สำหรับการเรียนรู้แบบจำลอง



ภาพที่ 2.8 เฟรมเวิร์คของ S-Sense **(Haruechaiyasak, Kongthon, Pornpimon, & Kanokorn, 2013)**

เฟรมเวิร์คของ S-Sense ดังแสดงในภาพที่ 2.8 นั้น ประกอบไปด้วยส่วนต่าง ๆ 4 ส่วนด้วยกัน ซึ่งสามารถอธิบายได้ดังนี้

1. **ส่วนของการเก็บรวบรวม และประมวลผลข้อความ (Text Collecting & Processing)** เป็นส่วนที่เกี่ยวข้องกับการไต่ (Crawling) และเก็บรวบรวม (Collecting) เนื้อหาจากเครือข่ายสังคมออนไลน์ ที่มาจากเว็บไซต์ต่าง ๆ ซึ่งในขั้นตอนนี้นั้น ได้รับถึงการประมวลผลข้อความพื้นฐาน อันประกอบไปด้วยการตัดประโยค (Sentence Segmentation) การทำให้เป็นโทเค่น (Tokenization) และการทำให้เป็นมาตรฐาน (Normalization) ที่หมายถึงการเปลี่ยนคำที่ปรากฏในข้อความ ให้อยู่ในรูปแบบปกติมาตรฐานทั่วไป อย่างคำว่า "Thnxsss" ก็จะถูกทำให้เป็นคำว่า "Thank" เป็นต้น
2. **ยูเรก้า (UREKA)** ซึ่งย่อมาจาก Utilization on REsource for Knowledge Acquisition เป็นขั้นตอนในการดึงลักษณะสำคัญที่เป็นใจหลักของคำ หรือ วลี ออกมาจากข้อความที่ได้รับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. **S-Sense** เป็นส่วนสำคัญหลักของเฟรมเวิร์คนี้ อันมีหน้าที่ในการวิเคราะห์ข้อมูล ซึ่ง S-Sense นั้น จะประกอบไปด้วย 4 โมดูลในการวิเคราะห์ด้วยกัน ได้แก่ 1) ส่วนการวิเคราะห์การใช้ภาษา (Language Usage Analysis) ที่ใช้จัดการกับภาษาที่ใช้แบบไม่เป็นทางการ ใช้ในการสนทนา หรือการใช้หยาบคาย (Obscene Language) ซึ่งการรับรู้ตรงสิ่งดังกล่าวนี้นั้นจะมีประโยชน์มาก เนื่องเพราะข้อความที่ให้อารมณ์ทางด้านขั้วลบนั้น มีเป็นจำนวนมาก ที่เป็นการใช้ภาษาที่หยาบคาย 2) ส่วนการวิเคราะห์ความตั้งใจ (Intention Analysis) ที่จะทำการแบ่งแต่ละข้อความออกเป็น 4 ประเภทด้วยกัน คือ การประกาศ (Announcement) การร้องขอ (Request) คำถาม (Question) และอารมณ์ (Sentiment) 3) ส่วนการวิเคราะห์อารมณ์ (Sentiment Analysis) สำหรับการวิเคราะห์แต่ละข้อความตามอารมณ์ของข้อความนั้น ซึ่งได้แก่ อารมณ์ทางขั้วบวก และอารมณ์ทางขั้วลบ 4) ส่วนการวิเคราะห์อารมณ์เชิงลึก (Emotion Analysis) ซึ่งเป็นงานในอนาคตที่จะทำการวิเคราะห์อารมณ์ลึกลงไป มากกว่าการวิเคราะห์เพียงขั้วอารมณ์ ว่าขั้วบวก หรือขั้วลบ เช่น อารมณ์เศร้า อารมณ์ดีใจ หรืออารมณ์กลัว เป็นต้น
4. **ส่วนเครื่องมือที่ใช้กำกับ และทรัพยากรทางภาษา (Tagging Tool and Language Resources)** ทรัพยากรทางภาษานั้น ประกอบไปด้วย 2 ส่วนประกอบด้วยกัน ได้แก่ คลังข้อมูลที่มีการอธิบายไว้ (Annotated Corpus) กับพจนานุกรมเจาะจงภาษา (Language-specific Lexicons) และเพื่อที่จะทำการสร้างทรัพยากรทางด้านภาษานั้น ก็ได้มีการจัดเตรียมเครื่องมือสำหรับการกำกับไว้ (Tagging Tool) สำหรับให้นักภาษาศาสตร์ได้ทำงานกับเครื่องมือดังกล่าว โดยเครื่องมือสำหรับการกำกับนั้น ถูกเขียนมาในรูปแบบเว็บแอพพลิเคชั่น (Web-based Application) ที่มีทั้งระบบจัดการฐานข้อมูล และส่วนประสานงานกับผู้ใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างระบบที่ขับเคลื่อนโดยใช้ S-Sense นั้น เช่น POP แสดงให้เห็นในภาพที่ 2.9 ซึ่งเป็นเว็บแอพพลิเคชั่นสำหรับการวิเคราะห์อารมณ์ ที่วิเคราะห์อารมณ์ข้อความที่ถูกทวีตออกมา จากทวิตเตอร์ (Twitter)

ภาพที่ 2.9 POP เว็บแอพพลิเคชั่นวิเคราะห์อารมณ์ ขับเคลื่อนโดย S-Sense

2.5 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่อง เป็นแขนงหนึ่งที่อยู่ในวิชาปัญญาประดิษฐ์ (Artificial Intelligence) ซึ่งเป็นศาสตร์ที่เกี่ยวข้องกับการจัดสร้าง และพัฒนา กระบวนการที่จะให้ระบบคอมพิวเตอร์นั้นสามารถทำการเรียนรู้สิ่งต่าง ๆ โดยวิเคราะห์จาก ข้อมูล พฤติกรรม หรือสิ่งที่มีอยู่แล้ว เพื่อนำสิ่งที่เรียนรู้นั้น มาเป็นสมมติฐานที่ใช้ในการจำแนกแยกแยะกลุ่มข้อมูลอื่น ๆ ต่อไป

โดยการเรียนรู้ของเครื่องนั้น หากแบ่งตามลักษณะของผลลัพธ์ที่ได้ มีประเภทหลัก ๆ ดังนี้

1. การเรียนรู้แบบมีผู้สอน (Supervised Learning)
2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)
3. การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning)
4. การเรียนรู้วิธีการเรียน (Learning to Learn, Meta-Learning)

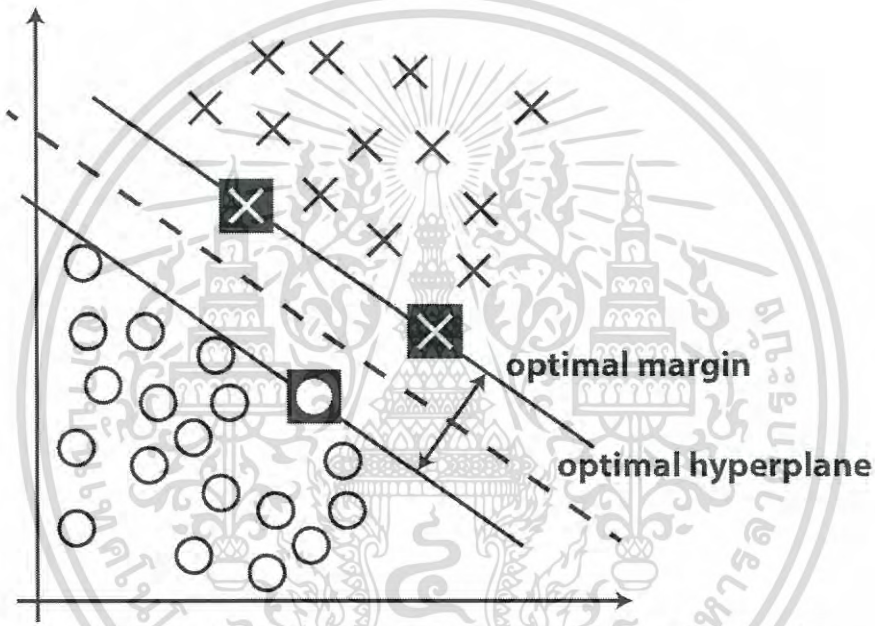
2.5.1 การเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised Machine Learning)

การเรียนรู้ของเครื่องแบบมีผู้สอน เป็นประเภทหนึ่งของการเรียนรู้ของเครื่อง ซึ่งอาศัยการเรียนรู้ จากกลุ่มข้อมูลสำหรับฝึกสอน (Training Set) ที่มีการดึงลักษณะสำคัญ (Feature) ออกมา และมีการกำหนดฉลาก (Label) ไว้ เพื่อให้ระบบสามารถเรียนรู้โดยการวิเคราะห์ข้อมูลจากลักษณะสำคัญ (Feature) และฉลาก (Label) ของกลุ่มข้อมูลสำหรับฝึกสอน (Training Set)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.2 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

ซัพพอร์ตเวกเตอร์แมชชีน หรือ เอสวีเอ็ม คือ แบบจำลองการเรียนรู้ของเครื่องแบบมีผู้สอนชนิดหนึ่ง ที่ริเริ่มโดย (Vapnik & Cortes, 1995) ซึ่งเดิมมีชื่อเรียกว่า ซัพพอร์ตเวกเตอร์เน็ตเวิร์คส์ (Support-Vector Networks) โดยเอสวีเอ็ม มีเป้าประสงค์ในการแก้ไขปัญหาการจำแนกข้อมูลแบบสองกลุ่ม (Two-group Classification Problems) โดยการนำข้อมูลรับเข้าที่เรียกว่า เวกเตอร์ของลักษณะสำคัญ (Feature Vector) ของข้อมูลแต่ละหน่วย มาแสดงเป็นจุด บนเวกเตอร์สเปซ (Vector Space) จากนั้นก็ทำการหาเส้นแบ่งข้อมูลทั้งสองประเภทออกจากกัน ซึ่งเส้นแบ่งนั้นเรียกว่า ไฮเปอร์เพลน (Hyperplane)



ภาพที่ 2.10 ตัวอย่างของซัพพอร์ตสเปซที่มีการแบ่งข้อมูลออกจากกันโดยเส้นไฮเปอร์เพลน **(Vapnik & Cortes, 1995)**

แรกเริ่ม เดิมทีนั้น อัลกอริทึมในการหาไฮเปอร์เพลน ที่ได้รับการเสนอโดย (Vapnik และ Cortes, 1995) นั้น จะเป็นอัลกอริทึมแบบเชิงเส้น กล่าวคือ เป็นการจำแนกประเภทข้อมูลแบบเชิงเส้น (Linear Classification) ต่อมา ได้มีการพัฒนาเคอร์เนลสำหรับการจำแนกประเภทข้อมูลแบบไม่เชิงเส้น (Nonlinear Classification) ขึ้น เช่น Polynomial หรือ RBF (Radial Basis Function) เป็นต้น

กรณีที่คลาส มีมากกว่าสองคลาสนั้น เพื่อที่จะสามารถกำหนดฉลาก (Label) ให้กับข้อมูลโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน ที่เดิมออกแบบมา สำหรับการแก้ปัญหาการจัดแบ่งประเภท แบบสองคลาส (Binary Classification Problem) นั้น ก็จำเป็นที่จะต้องมีเทคนิคมาช่วยในส่วนดังกล่าว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพื่อที่จะทำให้ ปัญหาการจัดแบ่งประเภทแบบหลายคลาส (Multiclass Classification Problem) กลายเป็นปัญหาการจัดแบ่งประเภทแบบสองคลาส (Binary Classification Problem) จำนวนหลาย ๆ ปัญหาแทน ซึ่งเทคนิคที่นิยมกันนั้น มีด้วยกันสองเทคนิค ได้แก่ One-versus-all ซึ่งเป็นการนำคลาสหนึ่ง มาเผชิญกับคลาสที่เหลือพร้อม ๆ กัน และ One-versus-one ซึ่งเป็นการนำคลาสหนึ่ง มาเผชิญกับคลาสอีกคลาสหนึ่ง

2.5.3 วิธีการตรวจสอบไขว้ (*K*-fold Cross Validation)

วิธีการตรวจสอบไขว้ (*K*-fold Cross Validation) เป็นเทคนิคในการวัดประสิทธิภาพการทำนายของแบบจำลอง โดยการสุ่มแบ่งข้อมูลออกเป็น K ส่วน และให้ K-1 ส่วน เป็นชุดข้อมูลสำหรับการฝึกสอน (Training Set) ส่วนอีก 1 ส่วน เป็นชุดข้อมูลสำหรับการทดสอบ (Testing Set) จากนั้นจึงทำการฝึกสอนด้วยชุดข้อมูลสำหรับฝึกสอน และทดสอบด้วยชุดข้อมูลสำหรับทดสอบ โดยทำเป็นจำนวน K ครั้ง โดยการสลับชุดข้อมูลสำหรับการทดสอบ (Testing Set) ไปทีละส่วน จนครบ K ส่วน (Kohavi, 1995)

2.5.4 ชุดข้อมูลที่ไม่สมดุล (Imbalanced Dataset)

ชุดข้อมูลที่ไม่สมดุล คือชุดข้อมูลที่จำนวนข้อมูลของแต่ละคลาสนั้น มีความต่างกันมาก ไม่มีความใกล้เคียงกัน อันปรากฏให้เห็นบ่อยครั้งในโลกของความเป็นจริง ที่จำนวนของข้อมูลลักษณะปกตินั้น มักจะมีมากกว่าจำนวนของข้อมูลที่มีลักษณะผิดปกติ โดยบางครั้งจำนวนของแต่ละคลาสอาจจะมีข้อมูลต่างกันถึง 100 เท่า เช่น กรณีที่พบได้บ่อย ๆ ในชุดข้อมูลเกี่ยวกับ การตรวจสอบทุจริต (Fraud Detection) หรือในบางครั้งจำนวนของแต่ละคลาส อาจจะมีข้อมูลต่างกันถึง 100,000 เท่า ซึ่งสามารถพบได้ในชุดข้อมูลประเภทอื่น ๆ (Chawla, Bowyer, Hall, และ Kegelmeyer, 2002)

การมีอยู่ของชุดข้อมูลที่ไม่สมดุลนั้น มักสร้างปัญหาให้กับการจำแนกประเภทของข้อมูล ที่ข้อมูลของคลาสที่มีข้อมูลจำนวนน้อยกว่า (Minority Class) นั้น ถูกจำแนกไปเป็นคลาสของข้อมูลที่มีจำนวนมากกว่า (Majority Class) ด้วยเหตุนี้ จึงมีการเสนอเทคนิค ในการจัดการกับชุดข้อมูลที่ไม่สมดุล เพื่อเพิ่มความแม่นยำของการจำแนกประเภทของข้อมูลให้มากขึ้น โดยต่อจากนี้จะกล่าวถึงเทคนิคต่าง ๆ ที่มีการใช้กัน

1. การลดจำนวนข้อมูลของคลาสที่มีจำนวนมาก (Majority Under-sampling) ให้มีจำนวนใกล้เคียงกับจำนวนข้อมูลของคลาสที่มีจำนวนน้อย (Minority Class) เป็นเทคนิคที่ลดจำนวนข้อมูลสำหรับการฝึกฝนของคลาสที่มีจำนวนมากออกไป ให้เหลือปริมาณใกล้เคียงกับคลาสที่มีจำนวนน้อย ซึ่งวิธีการนี้ ทำให้ต้องตัดข้อมูลที่มีคุณค่าออกไปโดยเปล่าประโยชน์
2. การเพิ่มจำนวนข้อมูลของคลาสที่มีจำนวนน้อย (Minority Over-sampling with Replacement) ให้มีจำนวนใกล้เคียงกับจำนวนข้อมูลของคลาสที่มีจำนวนมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Majority Class) โดยวิธีการคัดลอกข้อมูลเก่า คือ เทคนิคที่จะทำการคัดลอกข้อมูลของคลาสที่มีจำนวนน้อย ที่อยู่ในชุดข้อมูลสำหรับการฝึกสอนออกมา เพื่อเพิ่มจำนวนข้อมูลให้กับคลาสที่มีจำนวนน้อย ให้มีจำนวนใกล้เคียงกับคลาสที่มีจำนวนมาก โดยได้มีการถกเถียงไว้ใน (Nathalie, 2000) ซึ่งสรุปผลได้ว่า ไม่มีการเพิ่มประสิทธิภาพอย่างมีนัยยะสำคัญ ให้กับการตรวจจับข้อมูลของคลาสที่มีจำนวนน้อยเลย

3. การเพิ่มจำนวนข้อมูลของคลาสที่มีจำนวนน้อย โดยใช้เทคนิคการสังเคราะห์ข้อมูลขึ้นมา (Synthetic Minority Over-sampling Technique: SMOTE) ซึ่งได้รับการเสนอโดย (Chawla, Bowyer, Hall, & Kegelmeyer, 2002) โดยเทคนิคการสังเคราะห์ข้อมูลขึ้นมาของ SMOTE นั้น จะเป็นการสังเคราะห์ข้อมูลให้กับชุดข้อมูลสำหรับการฝึกฝนโดยการใช้การดำเนินการ (Operation) ที่แน่นอน กับชุดข้อมูลสำหรับการฝึกฝนที่เป็นข้อมูลจริง ๆ ที่มีอยู่ ซึ่งเป็นสิ่งที่ได้รับการพิสูจน์ว่า มีประสิทธิภาพสำหรับการนำไปใช้จำแนกข้อมูลมากกว่าการทำการลดจำนวนข้อมูลของคลาสที่มีจำนวนมาก (Under-sampling The Majority Class) อย่างมีนัยยะสำคัญ

2.6 ความแตกต่างระหว่างงานวิจัยนี้ และงานวิจัยที่เกี่ยวข้อง

งานวิจัย "การวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทยโดยใช้เซ็นติกเน็ต" กับงานวิจัยที่เกี่ยวข้องที่เป็นระบบวิเคราะห์อารมณ์ อันได้แก่ "ระบบวิเคราะห์อารมณ์ *pSenti*" และ "ระบบวิเคราะห์อารมณ์ S-Sense Social Sensing" มีข้อแตกต่างให้เห็นในหลายๆ แง่มุมด้วยกัน โดยสามารถแสดงให้เห็นความแตกต่างในดังตารางที่ 2.1

ตารางที่ 2.1 เปรียบเทียบระหว่างการวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทยโดยใช้เซ็นติกเน็ตและงานวิจัยที่เกี่ยวข้อง

งานวิจัย	ภาษา	ประเภทของการวิเคราะห์อารมณ์	Dataset	จำนวนคลาสในการแบ่งอารมณ์	Machine Learning Model
pSenti	อังกฤษ	Aspect Level	การรีวิวหนัง หรือสินค้า	คะแนนที่มีค่าระหว่าง -1 ถึง 1	SVM
S-Sense	ไทย	Sentence Level	ข้อความจากสื่อสังคมออนไลน์	Positive และ Negative	Naive Bayes
การวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทยโดยใช้เซ็นติกเน็ต	ไทย	Sentence Level	นิทานเด็ก 40 เรื่อง	Positive, Neutral และ Negative	SVM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยจากตารางแสดงให้เห็นความแตกต่างของ *pSenti* กับงานวิจัยอีกสองงาน ว่า *pSenti* นั้นเป็นงานวิจัยที่พัฒนาขึ้น สำหรับการวิเคราะห์ในภาษาอังกฤษ และเป็นแบบ Aspect Level ในขณะที่ S-Sense และการวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทยโดยใช้เซ็นติกเน็ต นั้นพัฒนาขึ้น สำหรับการวิเคราะห์อารมณ์ในภาษาไทย และเป็นการวิเคราะห์อารมณ์แบบ Sentence Level โดยงาน *pSenti* นั้น จะวิเคราะห์อารมณ์ออกมาเป็นค่าที่อยู่ระหว่าง -1 ถึง 1 ให้กับ Aspect แต่ละด้าน สำหรับงานวิจัย S-Sense จะเป็นการวิเคราะห์อารมณ์ภาพรวมของประโยค โดยแบ่งออกเป็นอารมณ์ทางด้านบวก กับอารมณ์ทางด้านลบ เท่านั้น ในขณะที่การวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทยโดยใช้เซ็นติกเน็ต ก็เป็นการวิเคราะห์อารมณ์ภาพรวมของประโยค เช่นกัน แต่ได้มีการแบ่งอารมณ์เป็น อารมณ์ทางด้านบวก อารมณ์ทางด้านลบ และอารมณ์กลาง

สำหรับส่วนการวิเคราะห์อารมณ์ของ S-Sense มีการแบ่งอารมณ์ออกเป็นสองคลาสนั้น และให้ค่าความแม่นยำ คือ 91.64% ส่วนการวิเคราะห์อารมณ์ของงานวิจัยนี้ มีการแบ่งอารมณ์ออกเป็นสามคลาส และให้ค่าความแม่นยำ 72.58% ด้วยกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
วิธีดำเนินการวิจัย

บทนี้จะเป็นการกล่าวถึงวิธีการในการดำเนินการวิจัย โดยเริ่มจากสมมติฐานการวิจัย ตลอดจนขั้นตอนต่าง ๆ ในการดำเนินการวิจัย ของ "การวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทยโดยใช้เซ็นติกเน็ต" อันเป็นงานวิจัยที่ได้มีการทำการรวบรวมฐานข้อมูลนิทานจำนวน 40 เรื่อง ที่มีการแบ่งประโยค และกำกับคลาสให้กับประโยคแต่ละประโยคไว้ ว่าประโยคดังกล่าวนั้น แสดงอารมณ์ขั้วใดออกมา ซึ่งแบ่งเป็น 3 อารมณ์ด้วยกัน ได้แก่ อารมณ์ด้านลบ อารมณ์ด้านบวก และอารมณ์กลาง หรือไม่แสดงอารมณ์ จากนั้นจึงมีการจัดสร้างทรัพยากรทางด้านอารมณ์ในภาษาไทย และมีการนำเสนอขั้นตอนต่าง ๆ ที่ทำให้ได้มาซึ่งเวกเตอร์ลักษณะสำคัญ โดยจะมีการนำการเรียนรู้ของเครื่องแบบมีผู้สอน มาใช้ในการจำแนกอารมณ์ของประโยคในนิทาน ออกเป็น 3 อารมณ์ดังที่ได้กล่าวไป จากนั้นจึงมีการวัดประสิทธิภาพของการทดลอง ด้วยวิธีการทางสถิติต่าง ๆ

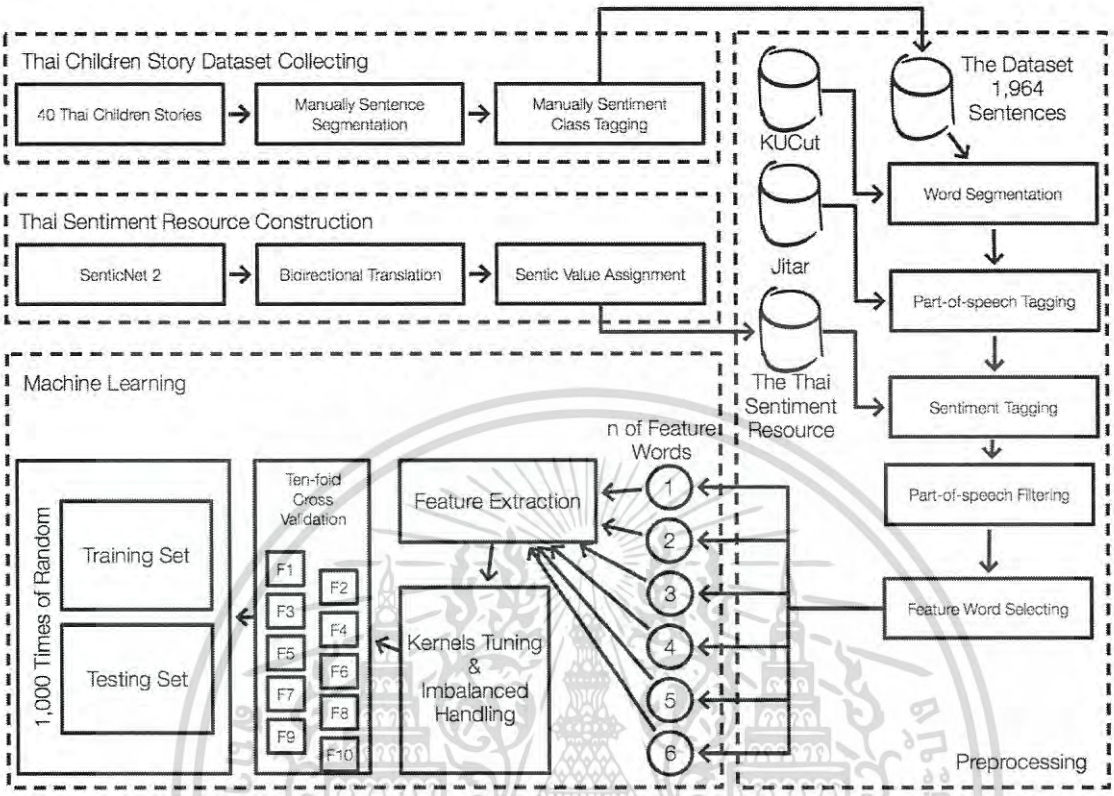
3.1 สมมติฐานการวิจัย

งานวิจัยนี้มีจุดประสงค์เพื่อนำเสนอกระบวนการในการวิเคราะห์อารมณ์ของประโยคข้อความภาษาไทย ที่อยู่ในนิทานเด็กจำนวน 40 เรื่อง โดยตั้งอยู่บนสมมติฐาน ดังนี้

1. อารมณ์ที่แต่ประโยคจะแสดงออกมานั้น สามารถเกิดขึ้นได้ 3 ประเภทด้วยกัน ได้แก่ อารมณ์ทางด้านบวก อารมณ์ทางด้านลบ และอารมณ์ที่เป็นกลาง
2. แต่ละประโยคนั้น จะแสดงอารมณ์ออกมาได้ เพียงแค่อารมณ์เดียวเท่านั้น จากอารมณ์ทั้ง 3 ประเภท
3. นิทานทั้ง 40 เรื่อง ที่ได้รับการกำกับอารมณ์ไว้โดยมนุษย์นั้น จะถือว่าเป็นการกำกับอารมณ์ที่ถูกต้อง และเป็นเกณฑ์ในการใช้วัดประสิทธิภาพ ของสิ่งที่งานวิจัยได้นำเสนอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 การรวบรวมชุดข้อมูลนิทาน 40 เรื่อง



ภาพที่ 3.1 กรอบแนวคิดการทำงานของการวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทยโดยใช้เซ็นติกเน็ต

ภาพที่ 3.1 กรอบสี่เหลี่ยมเส้นประด้านซ้ายบนสุดนั้น เป็นการแสดงกระบวนการในการรวบรวม ชุดข้อมูลนิทานทั้ง 40 เรื่อง ซึ่งกล่าวได้ ดังต่อไป

3.2.1 นิทาน 40 เรื่อง (40 Thai Children Stories)

นิทาน 40 เรื่องที่จะนำมาเป็นชุดข้อมูลจำหรับการวิจัยนี้นั้น ได้รับการรวบรวมใน (Netisopakul & Wikaha, 2014) คือนิทานสำหรับเด็ก ซึ่งประกอบไปด้วย เรื่องต่าง ๆ ดังนี้

1. หนูบ้านกับหนูนา
2. ลากับหมาป่า
3. ราชสีห์กบหนู
4. นกกระสากับหมาจิ้งจอก
5. ทำไมผึ้งจึงมีเหล็กใน
6. เต่าบินได้
7. ต้นอ้อผู้บอบบางกับต้นมะกอก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

8. คนตัดไม้ผู้ซื่อสัตย์
9. ต้นไม้แปลงร่าง
10. ห้องสมุดบินได้
11. สะพานลอยเกเรกับเจ้าด่าง
12. กระต่ายกับเต่า
13. พระอาทิตย์กับลมเหนือ
14. หมาป่ากับสุนัขบ้าน
15. เจ้าหญิงปุยฝ้ายกับเจ้าชายสายลม
16. กบผู้เฒ่านั่งเฝ้ากอบัว
17. ยายแห้งลอยฟ้า
18. แมงมุมใยทอง
19. ม้าแกลบน้อยของลีน่า
20. แพะอยากกินน้ำอัดลม
21. พรวิเศษจากนกน้อย
22. ผีน้อยลูแตง
23. หงส์ขนทอง
24. สร้อยคอนกเขา
25. ลิลลี่ทุ่งผู้แสนดี
26. ใบไม้วิเศษ
27. นิทานหิ่งห้อย
28. น้ำตาของเมฆน้อย
29. ห่านออกไข่เป็นทองคำ
30. ต้นไม้ร้องไห้
31. เสื้อกันหนาวสีแดงของลาน้อย
32. สุนัขจิ้งจอกกับนายพราน
33. แมลงหวี่น้อยผู้ยิ่งใหญ่
34. บุรุษไปรษณีย์ผู้น่ารัก
35. นกสาลิกากับนกยูง
36. ของขวัญของมาร์ติน
37. การผจญภัยของลูกค้างคาว
38. กระรอกน้อยกับนกเค้าแมว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

39. กระดิ่งผูกคอแมว

40. กบเลือกนาย

3.2.2 การตัดแบ่งประโยคโดยมนุษย์ (Manually Sentence Segmentation)

เนื่องจากการวิจัยนี้ เป็นการวิจัยการวิเคราะห์อารมณ์ ในระดับประโยค (Sentence-level Sentiment Analysis) อีกทั้งภาษาไทยนั้น ไม่มีสิ่งบ่งบอกถึงจุดสิ้นสุดของประโยคอย่างชัดเจน นิทานทั้ง 40 เรื่องนั้นจึงได้ถูกนำมาจัดแบ่งเป็นประโยคใหม่ โดยให้มนุษย์ทำการอ่านนิทานทั้งหมด และพิจารณาว่า จุดใด คือจุดที่สิ้นสุดของประโยคแต่ละประโยค จากนั้นจึงทำการแบ่งออกมาเป็นประโยคใหม่ จนกระทั่งสิ้นสุดนิทานทั้ง 40 เรื่อง ซึ่งนับจำนวนประโยคทั้งหมดได้ 1,964 ประโยค ด้วยกัน

3.2.3 การกำกับคลาสของอารมณ์โดยมนุษย์ (Manually Sentence Class Tagging)

แต่ละประโยคจากนิทานทั้ง 40 เรื่องที่ได้รับการตัดแบ่งแล้วนั้น จะได้รับการกำหนดค่าขั้วอารมณ์ขึ้น ซึ่งขั้วอารมณ์นั้น สามารถแบ่งได้เป็น ขั้วบวก ขั้วลบ และไม่มีขั้ว หรือค่ากลาง โดยผู้ทำการกำหนดค่าขั้วอารมณ์ให้กับประโยคแต่ละประโยค คือ มนุษย์ที่ทำการอ่านประโยคเหล่านั้น ซึ่งนิทานทั้ง 40 เรื่อง จำนวน 1,964 ประโยค หลังจากที่ได้รับการกำหนดค่าขั้วอารมณ์แล้วนั้น จึงประกอบไปด้วย

- ขั้วบวก 573 ประโยค หรือคิดเป็น 29.18%
- ขั้วลบ 451 ประโยค หรือคิดเป็น 22.96%
- ไม่มีขั้ว 940 ประโยค หรือคิดเป็น 47.86%

ในระหว่างที่มนุษย์ได้ทำการกำหนดอารมณ์ให้กับแต่ละประโยคนั้น หากมนุษย์ที่ทำการกำหนดอารมณ์ได้เห็นว่าในประโยคใด มีการปรากฏขึ้นของคำที่แสดงอารมณ์ออกมา ไม่ว่าจะเป็นทางบวก หรือทางลบ มนุษย์ก็จะทำการเลือก คำที่แสดงอารมณ์มากที่สุดตามที่มนุษย์ได้ลงความเห็นเอาไว้ด้วย และบันทึกคำดังกล่าวไว้ในระบบคู่กับประโยคดังกล่าว

สาเหตุที่ประโยคแต่ละประโยคในชุดข้อมูลนิทาน ต้องมีการกำหนดค่าขั้วของอารมณ์ให้นั้น เนื่องมาจากเป็นข้อกำหนดของการเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised Machine Learning) ที่จะต้องมีการสอนผลลัพธ์สุดท้ายให้กับเครื่อง เพื่อให้เครื่องนำ (Feature Vector) และผลลัพธ์สุดท้ายไปเรียนรู้ สร้างแบบจำลองออกมา ดังที่ได้กล่าวไว้ในหัวข้อที่ 2.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 การจัดสร้างทรัพยากรด้านอารมณ์ในภาษาไทย

ดังภาพที่ 3.1 นั้น กรอบสี่เหลี่ยมเส้นประอันที่สองทางด้านซ้ายมือ จะเป็นขั้นตอนการจัดสร้างทรัพยากรด้านอารมณ์ในภาษาไทย โดยประกอบไปด้วย 3 ส่วนหลัก ๆ ดังนี้

3.3.1 เซ็นติกเน็ต 2 (SenticNet 2)

เซ็นติกเน็ต 2 คือ ทรัพยากรทางด้านอารมณ์ที่ประกอบไปด้วยมโนทัศน์ภาษาอังกฤษ จำนวน 14,244 มโนทัศน์ อันประกอบไปด้วยมโนทัศน์ที่เป็นคำศัพท์ และมโนทัศน์ที่เป็นวลี ดังกล่าวไว้ในหัวข้อ 2.4.3 โดยเซ็นติกเน็ต 2 นั้น จะถูกนำมาใช้เป็นมโนทัศน์ตั้งต้น สำหรับการจัดสร้างทรัพยากรด้านอารมณ์ในภาษาไทยขึ้น ผ่านกระบวนการแปลแบบสองทาง

3.3.2 การแปลแบบสองทาง (Bidirectional Translation)

สำหรับมโนทัศน์จากเซ็นติกเน็ต 2 ทั้ง 14,244 มโนทัศน์ นั้น แต่ละคำมโนทัศน์ได้รับการแปลจากภาษาอังกฤษไปเป็นภาษาไทยโดยใช้พจนานุกรมอังกฤษ-ไทย LEXiTRON ของ NECTEC กับพจนานุกรมอังกฤษ-ไทย VOLUBILIS และแต่ละมโนทัศน์ในภาษาไทยที่เป็นผลลัพธ์จากการแปลจะถูกนำมาแปลกลับอีกครั้งให้เป็นภาษาอังกฤษ โดยใช้พจนานุกรมไทย-อังกฤษ LEXiTRON ของ NECTEC กับพจนานุกรมไทย-อังกฤษ VOLUBILIS ซึ่งหากมโนทัศน์ใดที่สามารถแปลกลับ และมีความหมายอันหนึ่ง อันใด ไปตรงกับมโนทัศน์ตั้งต้นแรกสุด มโนทัศน์ในภาษาไทยที่มีลักษณะเช่นนั้น ก็จะถูกเลือกมาเป็นผลลัพธ์ของการแปลแบบสองทาง ดังที่ได้อธิบายถึงการแปลแบบสองทางไว้ใน หัวข้อ 2.1.3

3.3.3 การกำหนดค่าอารมณ์ให้กับมโนทัศน์ที่ได้รับการแปล (Sentic Value Assignment)

สำหรับแต่ละมโนทัศน์ในภาษาไทยที่ได้รับการแปลมานั้น จะมีค่าเซ็นติกเหมือนมโนทัศน์ตั้งต้นทุกประการ ในกรณีที่มโนทัศน์ตั้งต้นของมโนทัศน์ที่ได้รับการแปลมานั้น มีเพียงมโนทัศน์เดียว ส่วนกรณีที่มโนทัศน์ตั้งต้นมีมากกว่าหนึ่งมโนทัศน์ แต่ละค่าของค่าเซ็นติก จะเป็นค่าเฉลี่ยของอารมณ์ด้านนั้น ๆ ที่มาจากค่าอารมณ์ของมโนทัศน์ตั้งต้นต่าง ๆ ทั้งหมด ซึ่งวิธีการเฉลี่ยค่านี้ คือสิ่งที่ได้ทำไปในเบื้องต้น ซึ่งอาจจะไม่ได้เป็นวิธีการที่ดีที่สุด เนื่องจากคำ ๆ หนึ่งนั้น อาจจะมีความหมายหลายนัย (Sense) โดยตัวอย่างทรัพยากรทางด้านอารมณ์ส่วนหนึ่งนั้น สามารถดูได้ที่ภาคผนวก

3.4 การเตรียมข้อมูล (Preprocessing)

ดังภาพที่ 3.1 นั้น กรอบสี่เหลี่ยมเส้นประทางด้านขวามือ จะแสดงให้เห็นถึงขั้นตอนการเตรียมข้อมูล ก่อนจะนำไปเข้าสู่ขั้นตอนการเรียนรู้ของเครื่อง โดยที่ชุดข้อมูลนิทาน 40 เรื่อง ที่ได้รับการรวบรวม และจัดสรรขึ้นมานั้น ได้มีการตัดแบ่งเป็นประโยคไว้ เป็นจำนวน 1,964 ประโยค และมี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การกำหนดค่าขั้วอารมณ์ไว้ ให้กับแต่ละประโยค ภายใน 3 คลาสด้วยกัน ขั้นตอนในหัวข้อนี้ จะเป็นการนำประโยคแต่ละประโยคนั้น มาเข้าสู่ขั้นตอนการเตรียมข้อมูล ซึ่งประกอบด้วยขั้นตอนต่าง ๆ ทางด้านการประมวลผลภาษาธรรมชาติ ดังที่จะกล่าวต่อไป

3.4.1 การตัดคำ (Word Segmentation)

สำหรับการตัดคำในประโยคแต่ละประโยคของนิทานภาษาไทยนั้น จะใช้เครื่องมือที่มีชื่อว่าคือ KUCut ในการตัดคำ โดยแต่ละประโยคในนิทาน จะผ่านขั้นตอนนี้ เป็นขั้นตอนแรกสุด เช่น ประโยคที่ว่า "มีหมาจิ้งจอกจอมเจ้าเล่ห์ตัวหนึ่งอาศัยอยู่ในป่าใกล้ทะเลสาบ มันขึ้นชื่อเรื่องการกลั่นแกล้งผู้อื่น" ก็จะได้รับการตัดคำออกมา และให้ผลลัพธ์เป็น "มี หมา จิ้งจอก จอม เจ้าเล่ห์ ตัว หนึ่ง อาศัย อยู่ ใน ป่า ใกล้ ทะเลสาบ _ มัน ขึ้นชื่อ เรื่อง การ กลั่นแกล้ง ผู้อื่น" เป็นต้น

3.4.2 การระบุชนิดของคำ (Part-of-speech Tagging)

ประโยคแต่ละประโยคที่ผ่านขั้นตอนการตัดคำมา จะมีคำศัพท์ที่แยกออกจากกันพร้อมเข้าสู่ขั้นตอนการระบุชนิดคำ โดยชนิดของคำที่ Jitar สามารถจะระบุได้ นั้น ประกอบไปด้วย กลุ่มชนิดของคำจำนวน 17 กลุ่ม และชนิดของคำจำนวน 48 ชนิด ที่แยกกันอยู่ตามกลุ่มต่าง ๆ ดังแสดงไว้ในสองคอลัมน์แรกของตารางที่ 3.1

ตัวอย่างการระบุชนิดของคำนั้น เช่น ประโยคที่ผ่านการตัดคำมาแล้ว ที่ว่า "มี หมา จิ้งจอก จอม เจ้าเล่ห์ ตัว หนึ่ง อาศัย อยู่ ใน ป่า ใกล้ ทะเลสาบ _ มัน ขึ้นชื่อ เรื่อง การ กลั่นแกล้ง ผู้อื่น" เมื่อผ่านเข้าขั้นตอนการระบุชนิดของคำ จะได้ผลลัพธ์ออกมาว่า "มี(*vt*) หมา(*ncn*) จิ้งจอก(*vt*) จอม(*ncn*) เจ้าเล่ห์(*ncn*) ตัว(*ncn*) หนึ่ง(*adj*) อาศัย(*vi*) อยู่(*vpost*) ใน(*prep*) ป่า(*ncn*) ใกล้(*prep*) ทะเลสาบ(*ncn*) _ (*blk*) มัน(*pper*) ขึ้นชื่อ(*vi*) เรื่อง(*ncn*) การ(*pref1*) กลั่นแกล้ง(*vt*) ผู้อื่น(*pind*)" เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 ชนิดของคำที่ Jitar ได้ทำการแบ่งไว้ และความถี่ของคำแสดงอารมณ์ที่มนุษย์เลือกมา

กลุ่มของชนิดของคำ	ชนิดของคำ	ความถี่
1. NOUN	1) *npn* (Proper noun)	20
	2) *nnum* (Cardinal number)	-
	3) *norm* (Ordinal Number Marker)	-
	4) *nlab* (Label noun)	-
	5) *ncn* (Common Noun)	73
	6) *nct* (Collective Noun)	-
	7) *ntit* (Title Noun)	-
2. PRONOUN	8) *pper* (Personal Pronoun)	-
	9) *pdem* (Demonstrative Pronoun)	-
	10) *pind* (Indefinite Pronoun)	-
	11) *ppos* (Possessive Pronoun)	-
	12) *prfx* (Reflexive Pronoun)	-
	13) *prec* (Reciprocal Pronoun)	-
	14) *prel* (Relative pronoun)	-
	15) *pint* (Interrogative Pronoun)	-
3. VERB	16) *vi* (Intransitive Verb)	518
	17) *vt* (Transitive Verb)	268
	18) *vcau* (Causative Verb)	-
	19) *vcs* (Complementary State Verb)	-
	20) *vex* (Existential Verb)	-
	21) *prev* (Pre-Verb)	1
	22) *vpost* (Post-verb)	-
	23) *honm* (Honorific marker)	-
4. DETERMINER	24) *det* (Determiner)	-
	25) *indet* (Indefinite determiner)	-
5. ADJECTIVE	26) *adj* (Adjective)	60
7. CLASSIFIER	33) *cl* (Classifier)	-
8. CONJUNCTION	34) *conj* (Conjunction)	1
	35) *conjd* (Double Conjunction)	-
	36) *conjncl* (Noun Clause Conjunction)	1
9. PREPOSITION	37) *prep* (Preposition)	-
	38) *prepc* (Co-Preposition)	-
10. INTERJECTION	39) *int* (Interjection)	1
11. PREFIX	40) *pref1* (Prefix1)	-
	41) *pref2* (Prefix2)	-
	42) *pref3* (Prefix3)	-
12. PARTICLE	43) *aff* (Affirmative)	2
	44) *part* (Particle)	1
13. NEGATIVE	45) *neg*	-
14. PUNCTUATION	46) *punc* (Punctuation)	-
15. IDIOM	47) *idm* (IDIOM)	1
16. PASSIVE VOICE MARKER	48) *psm* (Passive Voice Marker)	1
17. SYMBOL	49) *sym* (Symbol)	-

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.3 การระบุค่าอารมณ์ของคำ (Sentiment Tagging)

ทรัพยากรทางด้านอารมณ์ที่ถูกสร้างขึ้นจาก หัวข้อที่ 3.3 จะถูกนำมาใช้ทำการระบุค่าเซ็นติกให้กับคำศัพท์แต่ละคำในประโยค อันเป็นหนึ่งในขั้นตอนการเตรียมข้อมูลสำหรับการเรียนรู้ของเครื่องแบบมีผู้สอน เพื่อนำค่าเซ็นติกที่ระบุให้กับแต่ละคำนั้น เป็นส่วนหนึ่งในการดึงลักษณะสำคัญสำหรับการเรียนรู้ของเครื่องแบบมีผู้สอนต่อไป

ตัวอย่างการระบุค่าอารมณ์ของคำนั้น เช่น ประโยคที่ผ่านการตัดคำมาแล้วที่ว่า “มี หมา จิ้งจอก จอม เจ้าเล่ห์ ตัว หนึ่ง อาศัย อยู่ ใน ป่า ใกล้ ทะเลสาบ _ มัน ขึ้นชื่อ เรื่อง การ กลั่นแกล้ง ผู้อื่น” เมื่อเข้าสู่ขั้นตอนการระบุค่าอารมณ์ของคำ จะได้ผลลัพธ์ออกมาว่า

ตารางที่ 3.2 คำต่าง ๆ ในประโยค พร้อมค่าอารมณ์เซ็นติก

คำ	Pleasantness	Attention	Sensitivity	Aptitude	Polarity
มี	0.02	0.22	-0.17	0.29	0.11
หมา	0.0	0.0	0.02	-0.08	-0.03
จิ้งจอก	0.0	0.0	0.0	0.0	0.0
จอม	0.06	0.06	0.01	0.05	0.05
เจ้าเล่ห์	-0.38	0.17	-0.32	0.24	-0.1
ตัว	0.0	0.0	0.0	0.0	0.0
หนึ่ง	0.0	0.0	0.0	0.0	0.0
อาศัย	0.05	0.07	-0.04	0.11	0.07
อยู่	0.05	0.09	-0.07	0.13	0.07
ใน	0.0	0.1	0.0	0.12	0.07
ป่า	0.0	0.0	0.0	0.0	0.0
ใกล้	0.23	0.31	-0.14	0.34	0.25
ทะเลสาบ	0.0	0.0	0.0	0.0	0.0
_	0.0	0.0	0.0	0.0	0.0
มัน	0.01	0.07	-0.04	0.06	0.03
ขึ้นชื่อ	0.77	0.81	0.0	0.73	0.77
เรื่อง	0.03	0.35	-0.12	0.16	0.14
การ	0.17	0.19	-0.09	0.3	0.19
กลั่นแกล้ง	0.0	0.0	0.0	0.0	0.0
ผู้อื่น	0.01	0.31	-0.12	0.28	0.16

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.4 การคัดกรองคำ ตามประเภทของคำ (Part-of-speech Filtering)

หลังจากที่ประโยคแต่ละประโยค ได้รับการแบ่งเป็นหน่วยคำ พร้อมทั้งระบุชนิดของคำ และค่าอารมณ์แล้ว ขั้นตอนนี้ จะเป็นขั้นตอนการคัดกรองคำ เพื่อนเลือกคำที่เป็นผู้ท้าชิง (Candidate) ขึ้นมาสำหรับขั้นตอนการสกัดลักษณะสำคัญที่จะเกิดขึ้นต่อไป

ย้อนกลับไปในหัวข้อที่ 3.2.3 นั้น ระหว่างที่มนุษย์ได้ทำการกำกับคลาสของอารมณ์ให้กับแต่ละประโยค มนุษย์ก็ได้ทำการเลือกคำศัพท์ ที่มนุษย์เห็นว่าเป็นคำที่แสดงอารมณ์ออกมามากที่สุดไว้ด้วย ดังตารางที่ 3.1 นั้น จะเห็นได้ว่า ข้อมูลนิทานจำนวน 1,964 ประโยค มีประโยคที่แสดงค่าอารมณ์ไปทางบวก หรือทางลบ เป็นจำนวน 1,024 ประโยค ส่งผลให้มีคำที่แสดงอารมณ์ทั้งหมด 1,024 คำ ตามที่มนุษย์ได้ทำการเลือกคำแสดงอารมณ์สูงสุดไว้ จำนวน 1 คำ ต่อ 1 ประโยค โดยจากตารางนั้น สรุปได้ว่า

- มีจำนวนคำที่แสดงอารมณ์ที่มีชนิดของคำคือ *vi* (Intransitive Verb) จำนวน 518 คำ
- มีจำนวนคำที่แสดงอารมณ์ที่มีชนิดของคำคือ *vt* (Transitive Verb) จำนวน 268 คำ
- มีจำนวนคำที่แสดงอารมณ์ที่มีชนิดของคำคือ adv (Adverb) 76 คำ
- มีจำนวนคำที่แสดงอารมณ์ที่มีชนิดของคำคือ *ncn* (Common Noun) 73 คำ
- มีจำนวนคำที่แสดงอารมณ์ที่มีชนิดของคำคือ *adj* (Adjective) จำนวน 60 คำ
- มีจำนวนคำที่แสดงอารมณ์ที่มีชนิดของคำคือ *npn* (Proper noun) จำนวน 20 คำ
- มีจำนวนคำที่แสดงอารมณ์ที่มีชนิดของคำคือ *aff* (Affirmative) มีจำนวน 2 คำ

และสำหรับชนิดของคำ อันได้แก่ *prev* (Pre-Verb), *conj* (Conjunction), *conjncl* (Noun Clause Conjunction), *int* (Interjection), *part* (Particle), *idm* (IDIOM) และ *psm* (Passive Voice Marker) นั้น มีจำนวนความถี่ 1 เท่ากัน ส่วนชนิดของคำอื่น ๆ จากนี้ มีความถี่คือ 0 ซึ่ง หมายความว่า ไม่ปรากฏคำที่แสดงอารมณ์ที่ถูกจัดอยู่ในชนิดของคำดังที่กล่าวมา

โดยการคัดกรองคำในขั้นตอนนี้นั้น จะคัดกรองเฉพาะคำที่มี Part-of-speech เหมือนกับ 5 ลำดับต้นของ Part-of-speech ที่มีความถี่ของคำที่แสดงอารมณ์ออกมามากที่สุด ซึ่งได้แก่ *vi*, *vt*, adv, *ncn* และ *adj*

สำหรับตัวอย่างการคัดกรองคำ ตามประเภทของคำนั้น เช่น ประโยคที่ผ่านการตัดคำมาแล้ว ระบุชนิดของคำมาแล้ว และระบุค่าอารมณ์ของคำมาแล้ว ที่ว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.3 คำต่าง ๆ ในประโยค พร้อมค่าอารมณ์เซ็นติก และประเภทของคำ

คำ	POS	Pleasantness	Attention	Sensitivity	Aptitude	Polarity
มี	*vt*	0.02	0.22	-0.17	0.29	0.11
หมา	*ncn*	0.0	0.0	0.02	-0.08	-0.03
จิ้งจอก	*vt*	0.0	0.0	0.0	0.0	0.0
จอม	*ncn*	0.06	0.06	0.01	0.05	0.05
เจ้าเล่ห์	*ncn*	-0.38	0.17	-0.32	0.24	-0.1
ตัว	*ncn*	0.0	0.0	0.0	0.0	0.0
หนึ่ง	*adj*	0.0	0.0	0.0	0.0	0.0
อาศัย	*vi*	0.05	0.07	-0.04	0.11	0.07
อยู่	*vpost*	0.05	0.09	-0.07	0.13	0.07
ใน	*prep*	0.0	0.1	0.0	0.12	0.07
ป่า	*ncn*	0.0	0.0	0.0	0.0	0.0
ใกล้	*prep*	0.23	0.31	-0.14	0.34	0.25
ทะเลสาบ	*ncn*	0.0	0.0	0.0	0.0	0.0
_	*blk*	0.0	0.0	0.0	0.0	0.0
มัน	*pper*	0.01	0.07	-0.04	0.06	0.03
ขึ้นชื่อ	*vi*	0.77	0.81	0.0	0.73	0.77
เรื่อง	*ncn*	0.03	0.35	-0.12	0.16	0.14
การ	*prefl*	0.17	0.19	-0.09	0.3	0.19
กลั่นแกล้ง	*vt*	0.0	0.0	0.0	0.0	0.0
ผู้อื่น	*pind*	0.01	0.31	-0.12	0.28	0.16

เมื่อผ่านขั้นตอนการคัดกรองคำ ที่คัดเฉพาะคำที่มีประเภทของคำเป็น *vi*, *vt*, *adv*, *ncn* หรือ *adj* นั้น ก็จะเหลือเฉพาะคำเหล่านี้ เป็นผลลัพธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.4 คำต่าง ๆ ในประโยคที่มีประเภทของคำคือ *vi*, *vt*, *adv*, *ncn* หรือ *adj* พร้อมค่าอารมณ์เซ็นติก และประเภทของคำ

คำ	POS	Pleasantness	Attention	Sensitivity	Aptitude	Polarity
มี	*vt*	0.02	0.22	-0.17	0.29	0.11
หมา	*ncn*	0.0	0.0	0.02	-0.08	-0.03
จิ้งจอก	*vt*	0.0	0.0	0.0	0.0	0.0
จอม	*ncn*	0.06	0.06	0.01	0.05	0.05
เจ้าเล่ห์	*ncn*	-0.38	0.17	-0.32	0.24	-0.1
ตัว	*ncn*	0.0	0.0	0.0	0.0	0.0
หนึ่ง	*adj*	0.0	0.0	0.0	0.0	0.0
อาศัย	*vi*	0.05	0.07	-0.04	0.11	0.07
ป่า	*ncn*	0.0	0.0	0.0	0.0	0.0
ทะเลสาบ	*ncn*	0.0	0.0	0.0	0.0	0.0
ขึ้นชื่อ	*vi*	0.77	0.81	0.0	0.73	0.77
เรื่อง	*ncn*	0.03	0.35	-0.12	0.16	0.14
กลั่นแกล้ง	*vt*	0.0	0.0	0.0	0.0	0.0

จะสังเกตได้ว่า มีคำศัพท์หลายคำที่ไม่ควรมีค่าอารมณ์ แต่กลับมีค่าอารมณ์ เช่นคำว่า "การ" ในขณะที่คำที่ควรมีค่าอารมณ์อย่างคำว่า "กลั่นแกล้ง" กลับไม่มีค่าอารมณ์ออกมา ซึ่งนี่คือจุดหนึ่งที่แสดงให้เห็นถึงความไม่สมบูรณ์ มีข้อบกพร่อง ของทรัพยากรทางด้านอารมณ์ที่ได้เก็บรวบรวมไว้

3.4.5 การเลือกคำที่เป็นคำเด่นสำคัญ (Feature Word Selecting)

หลังจากผ่านขั้นตอนการตัดคำ การระบุชนิดของคำ การระบุค่าอารมณ์ของคำ และการคัดกรองคำ ตามประเภทของคำมาแล้วนั้น ขั้นตอนต่อไป จะเป็นขั้นตอนที่เลือกคำเด่นสำคัญ โดยเลือกจากคำที่เป็นผู้ท้าชิง (Candidate) ที่ยังคงหลงเหลืออยู่ ซึ่งเกณฑ์ในการเลือกคำนั้น จะทำการเลือกคำที่มีค่าขั้วของอารมณ์ (Polarity) ห่างจาก 0 มากที่สุด ออกมาเป็นเด่นสำคัญ เพื่อนำไปใช้ในขั้นตอนการเรียนรู้ของเครื่องต่อไป โดยคำที่มีค่าขั้วของอารมณ์ (Polarity) สูงสุด 6 อันดับแรกนั้น จะถูกเลือกออกมาเรียงลำดับกัน ในขั้นตอนนี้ เช่น ดังตารางที่ 3.4 นั้น เมื่อนำมาผ่านขั้นตอนนี้ ก็จะให้ผลลัพธ์ออกมาดัง ตารางที่ 3.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.5 คำที่มีค่า Polarity สูงสุด 6 อันดับแรกในประโยคที่มีประเภทของคำคือ *vi*, *vt*, adv, *ncn* หรือ *adj* พร้อมค่าอารมณ์เซ็นติก และประเภทของคำ

คำ	POS	Pleasantness	Attention	Sensitivity	Aptitude	Polarity
ขึ้นชื่อ	*vi*	0.77	0.81	0.0	0.73	0.77
เรื่อง	*ncn*	0.03	0.35	-0.12	0.16	0.14
มี	*vt*	0.02	0.22	-0.17	0.29	0.11
อาศัย	*vi*	0.05	0.07	-0.04	0.11	0.07
จอม	*ncn*	0.06	0.06	0.01	0.05	0.05
หมา	*ncn*	0.0	0.0	0.02	-0.08	-0.03

3.5 การเรียนรู้ของเครื่อง (Machine Learning)

3.5.1 การดึงลักษณะสำคัญสำหรับการเรียนรู้ของเครื่องแบบมีผู้สอน (Feature Extraction)

เมื่อประโยคแต่ละประโยคในนิทาน ได้ผ่านขั้นตอนการเตรียมข้อมูลจนเสร็จสิ้นแล้ว จนกระทั่งแต่ละประโยคนั้น ถูกเลือกคำออกมา 6 คำ ที่มีประเภทของคำเป็น *vi*, *vt*, *adv*, *ncn* หรือ *adj* อันเป็นกลุ่มคำที่มีค่าขั้วของอารมณ์ห่างจาก 0 มากที่สุด ขั้นตอนต่อไปนี้ จะเป็นขั้นการดึงลักษณะสำคัญออกมาจากประโยค โดยใช้คำเด่นสำคัญที่ถูกเลือกมาทั้ง 6 คำนั้น เพื่อการดึงลักษณะเด่นออกมา และจะใช้ค่าเซ็นติกของคำเหล่านั้น มาต่อกัน (Concatenation) ให้เป็นเวกเตอร์ลักษณะสำคัญสำหรับการเรียนรู้ของเครื่องต่อไป โดยสาเหตุที่นำเวกค่าเซ็นติกของแต่ละคำที่ถูกเลือกออกมา มาต่อกันนั้น เนื่องจากเห็นว่า ค่าอารมณ์ที่แสดงออกมาในภาพรวมของแต่ละประโยคนั้น เกิดจากหน่วยของคำ ที่อยู่ในประโยค และคำแต่ละหน่วยที่แสดงอารมณ์ออกมานั้น ไม่ได้มีความสัมพันธ์ต่อกัน หากแต่ต่างคำ ต่างมีลักษณะเด่นที่แก่งแย่ง เพื่อส่งผลกระทบต่ออารมณ์ภาพรวมของประโยค โดยเวกเตอร์ลักษณะสำคัญที่จะนำมาทำการทดลองนั้น จะประกอบด้วย 6 เวกเตอร์ด้วยกัน คือ เวกเตอร์ที่ประกอบด้วยคำที่มีค่าขั้วของอารมณ์ห่างจาก 0 มากที่สุด 1 อันดับแรก เวกเตอร์ที่ประกอบด้วยคำที่มีค่าขั้วของอารมณ์ห่างจาก 0 มากที่สุด 2 อันดับแรก 3 อันดับแรก 4 อันดับแรก 5 อันดับแรก และ 6 อันดับแรก ตัวอย่างเช่น คำเด่น 6 คำ ของประโยค “มีหมาจิ้งจอกจอมเจ้าเล่ห์ตัวหนึ่งอาศัยอยู่ในป่าใกล้ทะเลสาบ มันขึ้นชื่อเรื่องการกลั่นแกล้งผู้อื่น” ที่ได้แสดงไว้ในตารางที่ 3.5 ก็จะมีเวกเตอร์ของลักษณะสำคัญทั้ง 6 ดังนี้

- [0.77, 0.81, 0.0, 0.73, 0.77]
- [0.77, 0.81, 0.0, 0.73, 0.77, 0.03, 0.35, -0.12, 0.16, 0.14]
- [0.77, 0.81, 0.0, 0.73, 0.77, 0.03, 0.35, -0.12, 0.16, 0.14, 0.02, 0.22, -0.17, 0.29, 0.11]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [0.77, 0.81, 0.0, 0.73, 0.77, 0.03, 0.35, -0.12, 0.16, 0.14, 0.02, 0.22, -0.17, 0.29, 0.11, 0.05, 0.07, -0.04, 0.11, 0.07]
- [0.77, 0.81, 0.0, 0.73, 0.77, 0.03, 0.35, -0.12, 0.16, 0.14, 0.02, 0.22, -0.17, 0.29, 0.11, 0.05, 0.07, -0.04, 0.11, 0.07, 0.06, 0.06, 0.01, 0.05, 0.05]
- [0.77, 0.81, 0.0, 0.73, 0.77, 0.03, 0.35, -0.12, 0.16, 0.14, 0.02, 0.22, -0.17, 0.29, 0.11, 0.05, 0.07, -0.04, 0.11, 0.07, 0.06, 0.06, 0.01, 0.05, 0.05, 0.0, 0.0, 0.02, -0.08, -0.03]

3.5.2 การปรับเลือกเคอร์เนล และจัดการกับชุดข้อมูลที่ไม่สมดุล (Kernel Tuning & Imbalanced Dataset Handling)

จากขั้นตอนในหัวข้อ 3.5.1 ที่ผ่านมานั้น ส่งผลให้นิทานจำนวน 1,964 ประโยค มีเวกเตอร์ของลักษณะสำคัญ ประจำแต่ละประโยค ประโยคละ 6 เวกเตอร์ด้วยกัน ซึ่งการทดลองนั้น จะทำแยกกันสำหรับแต่ละเวกเตอร์ โดยแต่ละเวกเตอร์จะทำการทดลองโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน แบบ One-versus-all ด้วยเคอร์เนล 3 ประเภทด้วยกัน ได้แก่ Linear, Polynomial และ RBF และเนื่องจากลักษณะของข้อมูลนิทาน ที่จัดได้ว่าเป็นข้อมูลที่ไม่สมดุล (Imbalanced Dataset) เช่นนั้นแล้ว แต่ละเคอร์เนลจะมีการทำการทดลองด้วยกัน 4 การทดลองย่อยลงไปอีก โดยจะให้ 3 ใน 4 การทดลองหลังนั้น เป็นการใช้วิธีการ ในการจัดการกับข้อมูลที่ไม่สมดุล (Imbalanced Dataset) ซึ่งจะนำผลการทดลองทั้งหมดมาร่วมพิจารณาพร้อมกัน โดยการทดลองทั้ง 4 นั้น ได้แก่

1. การทดลองปกติที่ไม่มีการจัดการกับข้อมูลที่ไม่สมดุล
2. การทดลองที่มีการเปิดใช้งาน Class Weight
3. การทดลองที่มีการจัดการกับชุดข้อมูลที่ไม่สมดุล โดยใช้วิธีการเพิ่มจำนวนข้อมูลของคลาสที่มีจำนวนน้อย (Minority Over-sampling with Replacement) ซึ่งเป็นการสุ่มสร้างข้อมูลขึ้นมาซ้ำซ้อน จากข้อมูลเดิมที่มีอยู่ เพื่อให้ข้อมูลของแต่ละคลาส มีจำนวนเท่ากัน
4. การทดลองที่มีการจัดการกับชุดข้อมูลที่ไม่สมดุล โดยใช้วิธีการ Synthetic Minority Over-sampling Technique (SMOTE) ซึ่งเป็นการสร้างข้อมูลขึ้นมาเพิ่มให้กับคลาสที่มีจำนวนข้อมูลน้อย (Minority Class Over-sampling) โดยการสังเคราะห์ข้อมูลขึ้นมาดังที่ได้กล่าวไว้ใน หัวข้อที่ 2.5.4

ดังนั้นแล้ว การทดลองจึงมีทั้งหมด 72 การทดลองด้วยกัน คือการทดลองสำหรับเวกเตอร์ลักษณะสำคัญทั้ง 6 แบบ คูณกับเคอร์แนลทั้ง 3 แบบ และคูณกับการทดลองย่อยที่มีการจัดการกับชุดข้อมูลที่ไม่สมดุลย์อีก 4 แบบ ภายหลังจากทำการทดลองดังกล่าวแล้วนั้น จะนำผลลัพธ์ทั้งหมด 72 การทดลองมาเปรียบเทียบกัน ว่าการทดลองใด ให้ผลลัพธ์ดีที่สุด เพื่อหาเวกเตอร์ของลักษณะสำคัญที่เหมาะสมที่สุด เคอร์เนลที่เหมาะสมที่สุด และพารามิเตอร์ที่เหมาะสมที่สุด โดยการทดลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นั้น จะดำเนินไป โดยใช้ภาษา Python กับชุดคำสั่ง Scikit-learn (Cournapeau & Brucher, 2010) ในการทำการทดลอง

3.5.3 การตรวจสอบไขว้แบบ *K* ชุด (*K*-fold Cross Validation)

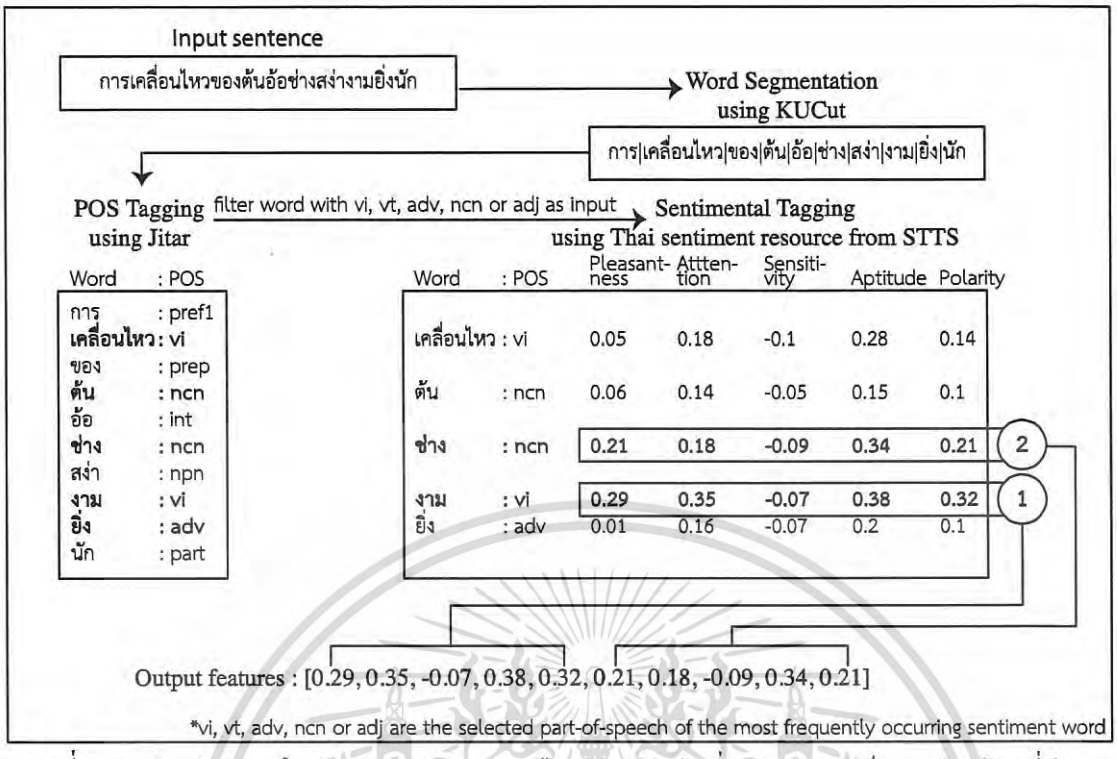
สำหรับการทดลองทั้ง 72 การทดลองนั้น จะมีการใช้ *K*-fold Cross Validation เพื่อทำการหาพารามิเตอร์ที่เหมาะสมที่สุดของการทดลองแต่ละการทดลอง โดยกำหนดค่า *K* ให้มีค่าเท่ากับ 10 ซึ่งพารามิเตอร์ดังกล่าว ที่จะถูกนำมาปรับค่า เพื่อหาชุดของพารามิเตอร์ที่ดีที่สุดนั้น ได้แก่ พารามิเตอร์ Regularisation (*C*) สำหรับทั้ง 3 เคอร์เนล พารามิเตอร์ *Degree* สำหรับเคอร์เนล Polymonial และพารามิเตอร์ *Gamma* สำหรับเคอร์เนล RBF โดยที่ค่าพารามิเตอร์ *C* จะถูกปรับอยู่ในช่วงระหว่าง 10^{-6} - 10^{4} ในขณะที่ค่าพารามิเตอร์ *Degree* จะถูกปรับอยู่ในช่วงระหว่าง 2 - 10 และค่าพารามิเตอร์ *Gamma* จะถูกปรับอยู่ในช่วงระหว่าง 10^{-6} - 10^{7} โดยชุดของพารามิเตอร์ที่ทำให้ผลลัพธ์ของ 10-fold Cross Validation ของแต่ละการทดสอบนั้น ออกมาดีที่สุด จะเป็นชุดของพารามิเตอร์ ที่ถูกนำไปใช้ในขั้นตอนถัดไป ในฐานะพารามิเตอร์ที่เหมาะสมที่สุด

3.5.4 การทำนายผลโดยใช้การเรียนรู้ของเครื่องแบบมีผู้สอน (Machine Learning Prediction)

หลังจากทำการหาค่าพารามิเตอร์ที่เหมาะสมที่สุดสำหรับแต่ละการทดลอง โดยการใช้ 10-fold Cross Validation แล้ว ประโยคจำนวน 1,964 ประโยคในนิทานนั้น จะได้รับการสุ่มเลือกขึ้นมา โดยแบ่งเป็นกลุ่มชุดข้อมูลสำหรับการฝึกฝน และกลุ่มชุดข้อมูลสำหรับการทดสอบ อย่างละครึ่ง ซึ่งก็คือกลุ่มละ 982 ประโยค และจะมีการแบ่งเช่นนี้เป็นจำนวน 1,000 ครั้งของการสุ่ม ด้วย Random Seed ที่แตกต่างกัน เพื่อลดความผิดพลาดทางสถิติ ที่เรียกว่า อคติที่เกิดจากการเลือกตัวอย่าง (Sampling Bias) ลง หลังจากนั้น ก็จะมีการสร้างแบบจำลองการเรียนรู้ของเครื่องแบบมีผู้สอนขึ้น โดยใช้ชุดข้อมูลสำหรับการฝึกฝน และพารามิเตอร์ที่เหมาะสมที่สุด ที่ได้รับมาจากขั้นตอน 10-fold Cross Validation เพื่อนำไปใช้ในการทำนายผลหาผลลัพธ์ทางสถิติต่อไป

เมื่อได้มีการนำชุดข้อมูลสำหรับฝึกสอน มาสร้างแบบจำลองสำหรับการเรียนรู้ของเครื่องขึ้น โดยทำการกำหนดค่าพารามิเตอร์ต่าง ๆ ตามพารามิเตอร์ที่เหมาะสมที่สุดสำหรับการทดลองนั้น อันได้รับมาจากการทำ 10-fold Cross Validation แล้ว ขั้นตอนถัดมา ก็จะเป็นการนำแบบจำลองดังกล่าวมาทำนายผลกับ กลุ่มชุดข้อมูลสำหรับการทดสอบ ที่อยู่ในชุดการทดลองเดียวกัน เพื่อทำการหาผลลัพธ์ทางสถิติของแต่ละชุดการทดลอง จากชุดการทดลองทั้ง 72 ชุด ชุดละ 1,000 ครั้งของการสุ่ม เพื่อนำมาเปรียบเทียบกันว่าการทดลองใด ให้ผลลัพธ์ดีที่สุด ด้วยกับเวกเตอร์ลักษณะสำคัญอันใด เคอร์เนลใด และพารามิเตอร์ใด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 3.2 ตัวอย่างการดึงลักษณะสำคัญ แบบเลือกสองคำเด่นที่แสดงอารมณ์ ของประโยคที่ว่า "การเคลื่อนไหวของต้นอ้อช่างสง่างามยิ่งนัก" **(Lertsuksakda, Pasupa, และ Netisopakul, Sentiment Analysis on Thai Children Stories with Support Vector Machine, 2015)**

สำหรับภาพที่ 3.2 เป็นการแสดงตัวอย่างขั้นตอนการดึงลักษณะสำคัญ ของประโยคที่ว่า "การเคลื่อนไหวของต้นอ้อช่างสง่างามยิ่งนัก" โดยประโยคดังกล่าวเริ่มต้นจากการเป็นข้อมูลนำเข้า ทางด้านบนซ้ายของภาพ และเข้าสู่ขั้นตอนการตัดคำออก จากนั้นแต่ละคำก็จะได้รับการกำหนดหน้าที่ของคำ จากนั้นจึงเข้าสู่ขั้นตอนการกำหนดค่าอารมณ์ของคำ โดยก่อนที่จะเข้าสู่ขั้นตอนนี้ เฉพาะคำศัพท์ที่มีหน้าที่ของคำเป็น *vi*, *vt*, *adv*, *ncn* หรือ *adj* เท่านั้น ที่จะถูกเลือกเข้าสู่ขั้นตอนดังกล่าว และผลลัพธ์ที่ได้ แสดงให้เห็นในกรอบสี่เหลี่ยมที่ใหญ่ที่สุดในภาพ ซึ่งคำที่มีค่า Polarity สูงที่สุดสองคำ จึงถูกเลือกออกมา โดยนำค่าอารมณ์ทั้ง 5 ค่า มาเรียงลำดับต่อกัน เพื่อเป็นเวกเตอร์ของลักษณะสำคัญ ที่จะใช้กับการเรียนรู้ของเครื่องต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการทดลอง

4.1 ผลการรวบรวมทรัพยากรด้านอารมณ์ในภาษาไทย

การรวบรวมทรัพยากรด้านอารมณ์ในภาษาไทย โดยใช้เซ็นติกเน็ต 2 (Cambria, Havasi, & Hussain, SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis, 2012) และใช้วิธีการแปลแบบสองทาง ซึ่งได้นำเสนอไว้ใน (Lertsuksakda, Netisopakul, และ Pasupa, Thai Sentiment Terms Construction using the Hourglass of Emotions, 2014) สามารถสรุปผลลัพธ์ได้ดังจะกล่าวต่อไป

ตารางที่ 4.1 ตัวอย่างมโนทัศน์ "โทสะ" ซึ่งถูกแปลแบบสองทางมาจาก 7 มโนทัศน์ในภาษาอังกฤษ

Concept\Sentic	Pleasantness	Attention	Sensitivity	Aptitude
Wrath	-0.737	+0.0	+0.899	+0.0
Hate	-0.714	+0.0	+0.83	-0.962
Hatred	-0.125	0.121	+0.446	-0.652
Fury	+0.0	+0.0	+0.786	-0.965
Rage	+0.0	+0.0	+0.859	+0.0
Animosity	-0.989	+0.0	+0.871	-0.976
Anger	+0.0	+0.0	+0.95	+0.0
ค่าเฉลี่ย	-0.366	0.017	+0.805	-0.507

การแปลมโนทัศน์จากเซ็นติกเน็ต 2 จำนวน 14,244 มโนทัศน์นั้น ปรากฏว่ามีมโนทัศน์ จำนวน 8,793 มโนทัศน์ ที่มีความหมายภาษาไทย ในพจนานุกรมอังกฤษ-ไทย LEXiTRON ของ NECTEC แต่เมื่อทำการแปลแบบสองทางโดยใช้ พจนานุกรมไทย-อังกฤษ LEXiTRON ในการแปลกลับแล้ว ปรากฏว่ามีมโนทัศน์เพียง 1,315 มโนทัศน์ ที่สามารถแปลไปกลับแบบสองทางได้ ซึ่งได้ผลลัพธ์เป็นมโนทัศน์ภาษาไทย จำนวน 1,440 มโนทัศน์ด้วยกัน และปรากฏว่ามีมโนทัศน์จำนวน 8,728 มโนทัศน์ ที่มีความหมายภาษาไทย ในพจนานุกรมอังกฤษ-ไทย VOLUBILIS แต่เมื่อทำการแปลแบบสองทางโดยใช้ พจนานุกรมไทย-อังกฤษ VOLUBILIS ในการแปลกลับแล้ว ปรากฏว่ามีมโนทัศน์จำนวน 4,688 มโนทัศน์ ที่สามารถแปลไปกลับแบบสองทางได้ ซึ่งได้ผลลัพธ์เป็นมโนทัศน์ภาษาไทย จำนวน 16,478 มโนทัศน์ด้วยกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการยูเนี่ยนระหว่างเซ็ตของคำแปลที่ได้มาจาก พจนานุกรมอังกฤษ-ไทย LEXiTRON ของ NECTEC จำนวน 1,440 มโนทัศน์ กับพจนานุกรมอังกฤษ-ไทย VOLUBILIS จำนวน 16,478 มโนทัศน์ ได้ให้ผลลัพธ์เป็นมโนทัศน์จำนวน 16,584 มโนทัศน์ด้วยกัน โดยสรุปผลได้ดังตารางที่ 4.2

ตารางที่ 4.2 สรุปผลการแปลเซ็นติกเน็ต 2 แบบสองทาง

พจนานุกรม	คำตั้งต้น	คำที่สามารถแปลได้	คำที่สามารถแปลแบบสองทางได้
LEXiTRON	14,244	8,793	1,315
VOLUBILIS	14,244	8,728	4,688
LEXiTRON และ VOLUBILIS	14,244	6,621	4,707

4.2 ผลการทดลองจากการเรียนรู้ของเครื่อง (Machine Learning)

การทดลองทั้ง 72 การทดลอง โดยใช้ซัพพอร์ตเวกเตอร์แมชชีน ดังได้กล่าวไว้ใน หัวข้อที่ 3.5 นั้น ภายหลังจากที่มีการดึงลักษณะสำคัญของแต่ละประโยคออกมา และปรับเลือกเคอร์แนล พร้อมจัดการกับข้อมูลที่ไม่สมดุลแล้ว ก็ได้มีการหาพารามิเตอร์ที่เหมาะสมที่สุด โดยใช้ Ten-fold Cross Validation ซึ่งสามารถสรุปผลจากการหาพารามิเตอร์ที่เหมาะสมที่สุด ได้ดัง ตารางที่ 4.3 ตารางที่ 4.4 ตารางที่ 4.5 และตารางที่ 4.6 ตามลำดับ โดยสังเกตได้จากคอลัมน์ที่ชื่อว่า "พารามิเตอร์"

หลังจากได้รับพารามิเตอร์ที่เหมาะสม สำหรับแต่ละชุดของการทดลองแล้วนั้น และได้ทำการแบ่งนิทานจำนวน 1,964 ประโยค เป็นชุดข้อมูลสำหรับการฝึกฝนครึ่งหนึ่ง 982 ประโยค และชุดข้อมูลสำหรับการทดสอบครึ่งหนึ่ง คือ 982 ประโยค เป็นจำนวน 1,000 ครั้ง ที่แตกต่างกัน จนกระทั่งทำการทดลองของการทดลองทั้ง 72 การทดลอง โดยแต่ละการทดลอง มีการทดลองย่อย 1,000 การทดลอง จนเสร็จสิ้นลง สามารถสรุปผลการทดลองเหล่านั้น ได้ดังนี้

4.2.1 ความแม่นยำ (Accuracy)

ค่าความแม่นยำ คือ เปอร์เซ็นความถูกต้องของการทำนายอารมณ์ของเครื่องซึ่งได้เรียนรู้ (Training) ชุดข้อมูลสำหรับการฝึกสอน (Training Set) ไปแล้ว ที่มาทำการทำนายข้อมูล ในชุดข้อมูลสำหรับการทดสอบ โดยค่าความแม่นยำ ณ ที่นี้ จะเป็นค่าเฉลี่ย ที่เกิดขึ้นจากการทดลองย่อยทั้ง 1,000 การทดลอง สำหรับการทดลองทั้ง 72 การทดลอง สามารถสรุปผลได้ดังดังตารางที่ 4.3 ตารางที่ 4.4 ตารางที่ 4.5 และตารางที่ 4.6 ตามลำดับ ซึ่งสามารถสรุปได้ว่า ค่าความแม่นยำของการทำนายผลนั้น จะมีผลลัพธ์ดีที่สุด โดยไม่ต้องการ Auto Class Weight หรือ Replacement หรือ SMOTE โดยที่สำหรับเคอร์เนล LINEAR นั้น หากมีจำนวนคำที่นำมาดึงลักษณะสำคัญจำนวน 3 คำ และพารามิเตอร์ $C = 1.0$ จะให้ผลลัพธ์ดีที่สุด ซึ่งมีความแม่นยำอยู่ที่ 67.96 ± 1.17 เปอร์เซ็น ส่วน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับเคอร์เนล POLYNOMIAL นั้น หากมีจำนวนคำที่นำมาดึงลักษณะสำคัญจำนวน 3 คำ และพารามิเตอร์ $C = 1.0$, $Degree = 2.0$ จะให้ผลลัพธ์ดีที่สุด ซึ่งมีความแม่นยำอยู่ที่ 65.81 ± 1.26 เปอร์เซ็น และสำหรับเคอร์เนล RBF นั้น หากมีจำนวนคำที่นำมาดึงลักษณะสำคัญจำนวน 2 คำ และพารามิเตอร์ $C = 1.0$, $Gamma = 0.1$ จะให้ผลลัพธ์ดีที่สุด ซึ่งมีความแม่นยำอยู่ที่ 72.58 ± 1.08 เปอร์เซ็น และเคอร์เนล RBF นั้น ได้ให้ผลลัพธ์ความแม่นยำในการทำนายมากที่สุด โดยมากกว่าเคอร์เนลอื่น ๆ ทั้งหมด

ตารางที่ 4.3 แสดงค่าความแม่นยำของการทำนายผลโดยใช้การเรียนรู้ของเครื่อง

เคอร์เนล	จำนวนคำที่นำมาดึงลักษณะสำคัญ	พารามิเตอร์	ค่าความแม่นยำ (%)
LINEAR	1	$C = 1.0$	65.46 ± 1.26
LINEAR	2	$C = 1.0$	67.42 ± 1.22
LINEAR	3	$C = 1.0$	67.96 ± 1.17
LINEAR	4	$C = 1.0$	67.69 ± 1.17
LINEAR	5	$C = 1.0$	67.42 ± 1.16
LINEAR	6	$C = 1.0$	67.27 ± 1.19
POLYNOMIAL	1	$C = 1.0$, $Degree = 2.0$	63.09 ± 1.16
POLYNOMIAL	2	$C = 1.0$, $Degree = 2.0$	65.81 ± 1.26
POLYNOMIAL	3	$C = 1.0$, $Degree = 2.0$	64.89 ± 1.23
POLYNOMIAL	4	$C = 1.0$, $Degree = 2.0$	63.53 ± 1.28
POLYNOMIAL	5	$C = 1.0$, $Degree = 2.0$	62.48 ± 1.26
POLYNOMIAL	6	$C = 1.0$, $Degree = 2.0$	61.97 ± 1.24
RBF	1	$C = 1.0$, $Gamma = 1.0$	67.79 ± 1.15
RBF	2	$C = 1.0$, $Gamma = 0.1$	72.58 ± 1.08
RBF	3	$C = 1.0$, $Gamma = 0.1$	71.45 ± 1.08
RBF	4	$C = 1.0$, $Gamma = 0.1$	70.50 ± 1.12
RBF	5	$C = 1.0$, $Gamma = 0.01$	69.01 ± 1.12
RBF	6	$C = 1.0$, $Gamma = 0.001$	69.01 ± 1.14

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.4 แสดงค่าความแม่นยำของการทำนายผลโดยใช้การเรียนรู้ของเครื่องแบบมี Auto Class Weight

เคอร์เนล	จำนวนคำที่นำมาดึงลักษณะสำคัญ	พารามิเตอร์	ค่าความแม่นยำ (%)
LINEAR	1	*C* = 10.0	64.38 ± 1.25
LINEAR	2	*C* = 1.0	66.43 ± 1.19
LINEAR	3	*C* = 1.0	66.51 ± 1.24
LINEAR	4	*C* = 1.0	65.83 ± 1.28
LINEAR	5	*C* = 1.0	65.35 ± 1.32
LINEAR	6	*C* = 1.0	64.79 ± 1.30
POLYNOMIAL	1	*C* = 10.0, *Degree* = 2.0	63.48 ± 1.16
POLYNOMIAL	2	*C* = 1.0, *Degree* = 2.0	67.42 ± 1.15
POLYNOMIAL	3	*C* = 1.0, *Degree* = 2.0	66.69 ± 1.12
POLYNOMIAL	4	*C* = 1.0, *Degree* = 2.0	65.56 ± 1.10
POLYNOMIAL	5	*C* = 1.0, *Degree* = 2.0	64.45 ± 1.11
POLYNOMIAL	6	*C* = 1.0, *Degree* = 2.0	63.84 ± 1.10
RBF	1	*C* = 1000.0, *Gamma* = 0.01	66.17 ± 1.29
RBF	2	*C* = 100.0, *Gamma* = 0.01	70.62 ± 1.21
RBF	3	*C* = 1000.0, *Gamma* = 0.001	70.26 ± 1.10
RBF	4	*C* = 10.0, *Gamma* = 0.01	69.81 ± 1.13
RBF	5	*C* = 1.0, *Gamma* = 0.1	68.26 ± 1.24
RBF	6	*C* = 10.0, *Gamma* = 0.01	68.41 ± 1.15

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 แสดงค่าความแม่นยำของการทำนายผลโดยใช้การเรียนรู้ของเครื่องแบบมีการ

Replacement

เคอร์เนล	จำนวนคำที่นำมาดึงลักษณะสำคัญ	พารามิเตอร์	ค่าความแม่นยำ (%)
LINEAR	1	$C = 1.0$	64.33 ± 1.27
LINEAR	2	$C = 1.0$	66.43 ± 1.20
LINEAR	3	$C = 1.0$	66.48 ± 1.27
LINEAR	4	$C = 1.0$	65.79 ± 1.29
LINEAR	5	$C = 1.0$	65.28 ± 1.35
LINEAR	6	$C = 1.0$	64.72 ± 1.36
POLYNOMIAL	1	$C = 1.0$, *Degree* = 2.0	63.38 ± 1.18
POLYNOMIAL	2	$C = 1.0$, *Degree* = 2.0	67.76 ± 1.18
POLYNOMIAL	3	$C = 1.0$, *Degree* = 2.0	66.79 ± 1.12
POLYNOMIAL	4	$C = 1.0$, *Degree* = 2.0	65.55 ± 1.15
POLYNOMIAL	5	$C = 1.0$, *Degree* = 2.0	64.36 ± 1.12
POLYNOMIAL	6	$C = 1.0$, *Degree* = 2.0	63.72 ± 1.13
RBF	1	$C = 1.0$, *Gamma* = 1.0	66.19 ± 1.24
RBF	2	$C = 1.0$, *Gamma* = 0.1	71.17 ± 1.24
RBF	3	$C = 10.0$ *Gamma* = 0.01	70.37 ± 1.13
RBF	4	$C = 1.0$, *Gamma* = 0.01	69.22 ± 1.14
RBF	5	$C = 100.0$, *Gamma* = 0.001	68.78 ± 1.17
RBF	6	$C = 100.0$, *Gamma* = 0.001	68.44 ± 1.15

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.6 แสดงค่าความแม่นยำของการทำนายผลโดยใช้การเรียนรู้ของเครื่องแบบมี SMOTE

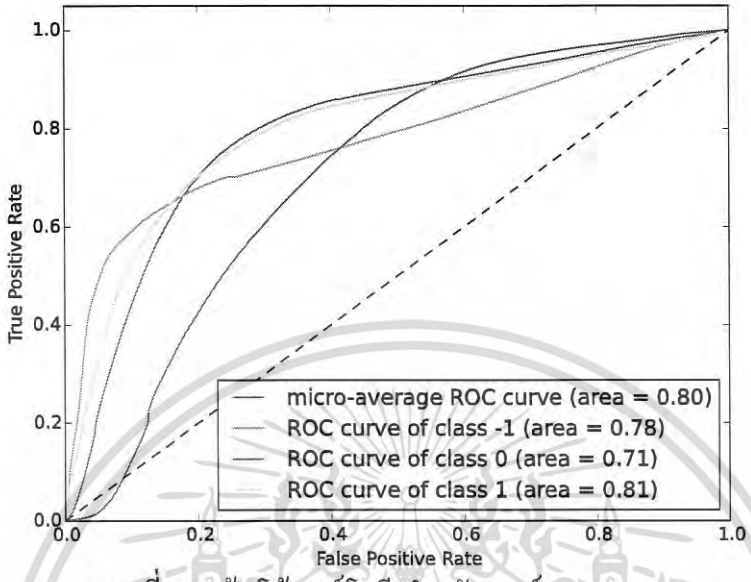
เคอร์เนล	จำนวนคำที่นำมาดึงลักษณะสำคัญ	พารามิเตอร์	ค่าความแม่นยำ (%)
LINEAR	1	$C = 1.0$	64.52 ± 1.23
LINEAR	2	$C = 1.0$	66.48 ± 1.21
LINEAR	3	$C = 1.0$	66.40 ± 1.27
LINEAR	4	$C = 1.0$	65.71 ± 1.29
LINEAR	5	$C = 1.0$	65.26 ± 1.33
LINEAR	6	$C = 1.0$	64.69 ± 1.32
POLYNOMIAL	1	$C = 1.0$ *Degree* = 6.0	60.31 ± 1.20
POLYNOMIAL	2	$C = 1.0$ *Degree* = 10.0	61.30 ± 1.24
POLYNOMIAL	3	$C = 1.0$ *Degree* = 8.0	60.17 ± 1.18
POLYNOMIAL	4	$C = 1.0$ *Degree* = 10.0	57.56 ± 1.25
POLYNOMIAL	5	$C = 1.0$ *Degree* = 10.0	56.47 ± 1.30
POLYNOMIAL	6	$C = 1.0$ *Degree* = 9.0	56.60 ± 1.26
RBF	1	$C = 1.0$ *Gamma* = 1.0	66.15 ± 1.23
RBF	2	$C = 1.0$ *Gamma* = 0.1	70.96 ± 1.25
RBF	3	$C = 1000.0$ *Gamma* = 0.001	69.64 ± 1.18
RBF	4	$C = 1.0$ *Gamma* = 0.1	68.69 ± 1.23
RBF	5	$C = 100.0$ *Gamma* = 0.001	68.13 ± 1.19
RBF	6	$C = 100.0$ *Gamma* = 0.001	67.75 ± 1.18

4.2.2 เส้นโค้งอาร์โอซี (Receiver Operating Characteristic Curve: ROC Curve)

เส้นโค้งอาร์โอซี เป็นการแสดงผลในรูปแบบกราฟ เพื่อแสดงให้เห็นประสิทธิภาพของผลลัพธ์การจำแนกประเภท (Classifier Output Quality) ซึ่งเป็นการแสดงความสัมพันธ์ระหว่าง True Positive Rate (หรือ Sensitivity) กับ False Positive Rate (หรือ 1 - Specificity) สำหรับค่า Classification Threshold ที่เป็นไปได้ทั้งหมด ส่วนพื้นที่ใต้กราฟ (AUC: Area Under The Curve) นั้น จะแสดงให้เห็นถึงประสิทธิภาพของตัวจำแนกประเภท (Classfier) ที่หากยิ่งมีค่ามาก ยิ่งบ่งบอกถึงความมีประสิทธิภาพ ในขณะที่หากมีค่าพื้นที่ใต้กราฟน้อยใกล้กับ 0.5 ก็จะเป็นการบ่งบอกว่าตัวจำแนกประเภทนั้น แย่ หรือไม่มีประสิทธิภาพนั่นเอง โดยการแสดงเส้นโค้งอาร์โอซีนั้นจะแสดงเป็นจำนวน 3 กราฟด้วยกัน ซึ่งเป็นกราฟสำหรับผลลัพธ์ที่ดีที่สุด ของเคอร์เนลทั้ง 3 อัน ได้แก่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

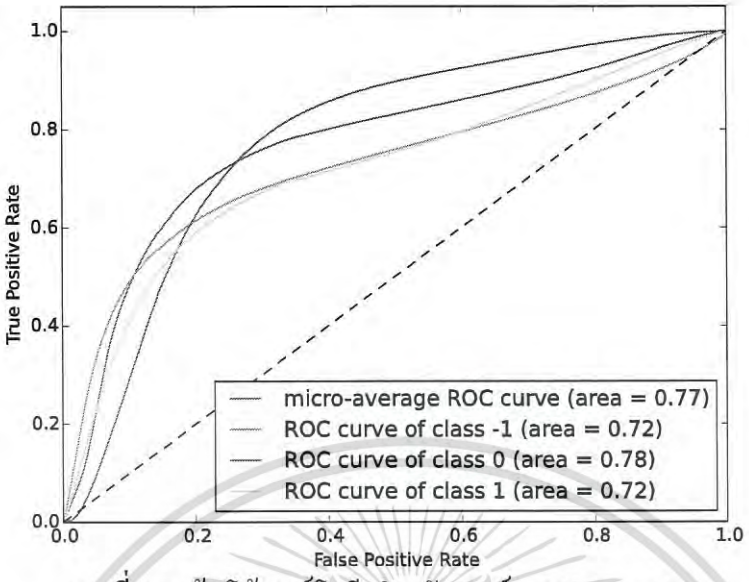
LINEAR, POLYNOMIAL และ RBF ดังที่ได้กล่าวไว้ในหัวข้อที่ 4.2.1 ซึ่งกราฟสามารถแสดงได้ดังภาพที่ 4.1 ภาพที่ 4.2 และภาพที่ 4.3 ตามลำดับ



ภาพที่ 4.1 เส้นโค้งอาร์โอซี สำหรับเคอร์เนล Linear

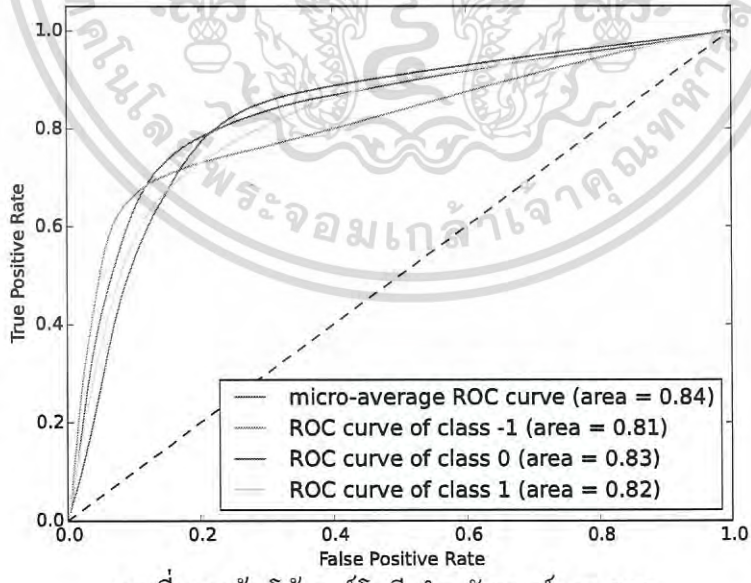
ภาพที่ 4.1 เส้นโค้งอาร์โอซี ประกอบไปด้วยเส้นโค้งจำนวน 4 เส้น ได้แก่ เส้นโค้งของคลาสขั้วลบ (Negative Class) ที่มีพื้นที่ใต้กราฟคือ 0.78 เส้นโค้งของคลาสขั้วกลาง (Neutral Class) ที่มีพื้นที่ใต้กราฟคือ 0.71 เส้นโค้งของคลาสขั้วบวก (Positive Class) ที่มีพื้นที่ใต้กราฟคือ 0.81 และสุดท้าย คือ เส้นโค้งค่าเฉลี่ยที่มีพื้นที่ใต้กราฟ 0.80

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 4.2 เส้นโค้งอาร์โอซี สำหรับเคอร์เนล Polynomial

ภาพที่ 4.2 เส้นโค้งอาร์โอซี ประกอบไปด้วยเส้นโค้งจำนวน 4 เส้น ได้แก่ เส้นโค้งของคลาสขั้วลบ (Negative Class) ที่มีพื้นที่ใต้กราฟคือ 0.72 เส้นโค้งของคลาสขั้วกลาง (Neutral Class) ที่มีพื้นที่ใต้กราฟคือ 0.78 เส้นโค้งของคลาสขั้วบวก (Positive Class) ที่มีพื้นที่ใต้กราฟคือ 0.72 และสุดท้าย คือ เส้นโค้งค่าเฉลี่ยที่มีพื้นที่ใต้กราฟ 0.77



ภาพที่ 4.3 เส้นโค้งอาร์โอซี สำหรับเคอร์เนล RBF

ภาพที่ 4.3 เส้นโค้งอาร์โอซี ประกอบไปด้วยเส้นโค้งจำนวน 4 เส้น ได้แก่ เส้นโค้งของคลาสขั้วลบ (Negative Class) ที่มีพื้นที่ใต้กราฟคือ 0.81 เส้นโค้งของคลาสขั้วกลาง (Neutral Class)

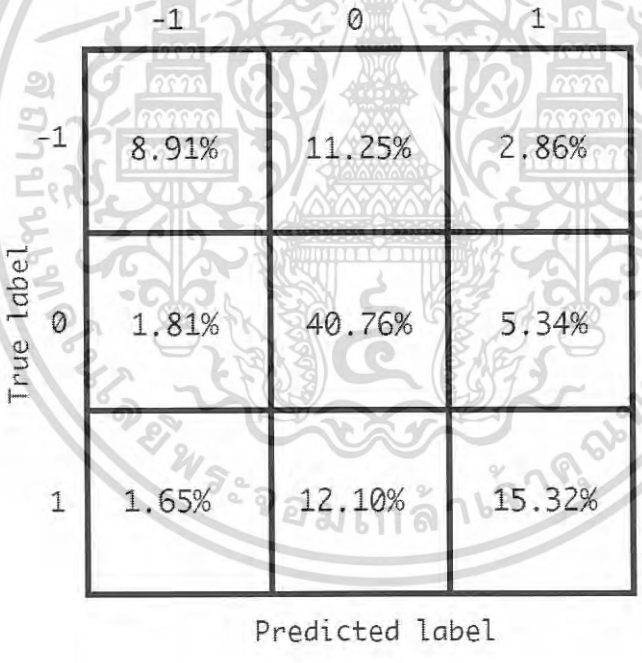
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ที่มีพื้นที่ใต้กราฟคือ 0.83 เส้นโค้งของคลาสขั้วบวก (Positive Class) ที่มีพื้นที่ใต้กราฟคือ 0.82 และสุดท้าย คือ เส้นโค้งค่าเฉลี่ยที่มีพื้นที่ใต้กราฟ 0.84

4.2.3 เมตริกซ์วัดประสิทธิภาพ (Confusion Matrix)

เมตริกซ์วัดประสิทธิภาพ (Confusion Matrix) สำหรับวัดประสิทธิภาพผลลัพธ์จากการจำแนกประเภท (Classifier Output Quality) โดยค่าที่ได้รับการทำนายอย่างถูกต้องนั้น จะแสดงอยู่ในแกนเส้นทแยงมุมลากจากมุมซ้ายบน ไปยังมุมขวาล่าง ในขณะที่ค่าที่ได้รับการทำนายและผิด จะแสดงอยู่ในช่องสี่เหลี่ยม 6 ช่องรอบด้านที่เหลือ ซึ่งสำหรับแกนวาย (Y-axis) นั้น คือค่าผลลัพธ์ในความเป็นจริงที่สมควรเป็น ส่วนแกนเอ็กซ์ (X-axis) นั้น คือค่าที่ได้จากการทำนาย

เมตริกซ์วัดประสิทธิภาพนี้ จะแสดงเป็นจำนวน 3 เมตริกซ์ด้วยกัน อันเป็นเมตริกซ์สำหรับผลลัพธ์ที่ดีที่สุด ของเคอร์เนลทั้ง 3 อัน ได้แก่ LINEAR, POLYNOMIAL และ RBF ดังที่ได้กล่าวไว้ในหัวข้อที่ 4.2.1 ซึ่งเมตริกซ์ทั้ง 3 จะสามารถแสดงได้ดังภาพที่ 4.4 ภาพที่ 4.5 และภาพที่ 4.6

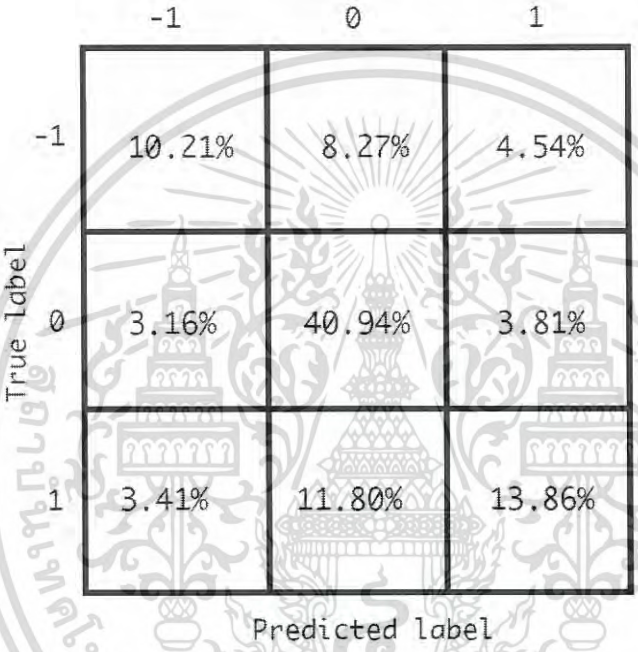


ภาพที่ 4.4 เมตริกซ์วัดประสิทธิภาพ สำหรับเคอร์เนล Linear

ภาพที่ 4.4 แสดงความสัมพันธ์ระหว่างค่าอารมณ์จริง กับค่าอารมณ์ที่ได้รับการทำนาย ของชุดสำหรับการทดสอบ (Testing Set) จากการทดลองย่อยทั้ง 1,000 ครั้ง ซึ่งอธิบายได้ว่า คลาสขั้วลบ (Negative Class) นั้น มีจำนวน 23.02% ซึ่งได้รับการทำนายอย่างถูกต้อง เป็นคลาสขั้วลบ (Negative Class) 8.91% ได้รับการทำนายผิดเป็นคลาสขั้วกลาง (Neutral Class) 11.25% และได้รับการทำนาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผิดเป็นคลาสขั้วบวก (Positive Class) 2.86% ในขณะที่คลาสขั้วกลาง (Neutral Class) นั้น มีจำนวน 47.91% ซึ่งได้รับการทำนายอย่างถูกต้อง เป็นคลาสขั้วกลาง (Neutral Class) 40.76% ได้รับการทำนายผิดเป็นคลาสขั้วลบ (Negative Class) 1.81% และได้รับการทำนายผิดเป็นคลาสขั้วบวก 5.34% และในขณะที่คลาสขั้วบวก (Positive Class) นั้น มีจำนวน 29.07% ซึ่งได้รับการทำนายอย่างถูกต้อง เป็นคลาสขั้วบวก (Positive Class) 15.32% ได้รับการทำนายผิดเป็นคลาสขั้วลบ (Negative Class) 1.65% และได้รับการทำนายผิดเป็นคลาสขั้วกลาง (Neutral Class) 12.10%

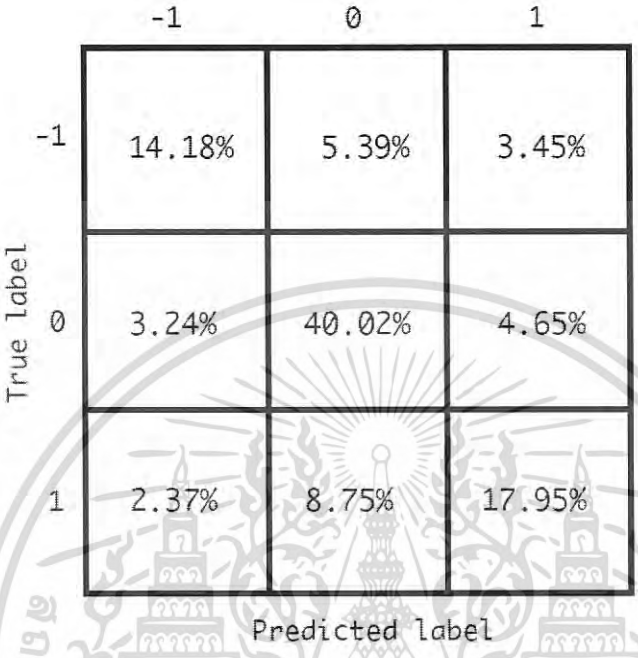


ภาพที่ 4.5 เมตริกซ์วัดประสิทธิภาพ สำหรับเคอร์เนล Polynomial

ภาพที่ 4.5 แสดงความสัมพันธ์ระหว่างค่าอารมณ์จริง กับค่าอารมณ์ที่ได้รับการทำนาย ของชุดสำหรับการทดสอบ (Testing Set) จากการทดลองย่อยทั้ง 1,000 ครั้ง ซึ่งอธิบายได้ว่า คลาสขั้วลบ (Negative Class) นั้น มีจำนวน 23.02% ซึ่งได้รับการทำนายอย่างถูกต้อง เป็นคลาสขั้วลบ (Negative Class) 10.21% ได้รับการทำนายผิดเป็นคลาสขั้วกลาง (Neutral Class) 8.27% และได้รับการทำนายผิดเป็นคลาสขั้วบวก (Positive Class) 4.54% ในขณะที่คลาสขั้วกลาง (Neutral Class) นั้น มีจำนวน 47.91% ซึ่งได้รับการทำนายอย่างถูกต้อง เป็นคลาสขั้วกลาง (Neutral Class) 40.94% ได้รับการทำนายผิดเป็นคลาสขั้วลบ (Negative Class) 3.16% และได้รับการทำนายผิดเป็นคลาสขั้วบวก 3.81% และในขณะที่คลาสขั้วบวก (Positive Class) นั้น มีจำนวน 29.07% ซึ่งได้รับการทำนายอย่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถูกต้อง เป็นคลาสขั้วบวก (Positive Class) 13.86% ได้รับการทำนายผิดเป็นคลาสขั้วลบ (Negative Class) 3.41% และได้รับการทำนายผิดเป็นคลาสขั้วกลาง (Neutral Class) 11.80%



ภาพที่ 4.6 เมตริกซ์วัดประสิทธิภาพ สำหรับเคอร์เนล RBF

ภาพที่ 4.6 แสดงความสัมพันธ์ระหว่างค่าอารมณ์จริง กับค่าอารมณ์ที่ได้รับการทำนาย ของชุดสำหรับการทดสอบ (Testing Set) จากการทดลองย่อยทั้ง 1,000 ครั้ง ซึ่งอธิบายได้ว่า คลาสขั้วลบ (Negative Class) นั้น มีจำนวน 23.02% ซึ่งได้รับการทำนายอย่างถูกต้อง เป็นคลาสขั้วลบ (Negative Class) 14.18% ได้รับการทำนายผิดเป็นคลาสขั้วกลาง (Neutral Class) 5.39% และได้รับการทำนายผิดเป็นคลาสขั้วบวก (Positive Class) 3.45% ในขณะที่คลาสขั้วกลาง (Neutral Class) นั้น มีจำนวน 47.91% ซึ่งได้รับการทำนายอย่างถูกต้อง เป็นคลาสขั้วกลาง (Neutral Class) 40.02% ได้รับการทำนายผิดเป็นคลาสขั้วลบ (Negative Class) 3.24% และได้รับการทำนายผิดเป็นคลาสขั้วบวก 4.65% และในขณะที่คลาสขั้วบวก (Positive Class) นั้น มีจำนวน 29.07% ซึ่งได้รับการทำนายอย่างถูกต้อง เป็นคลาสขั้วบวก (Positive Class) 17.95% ได้รับการทำนายผิดเป็นคลาสขั้วลบ (Negative Class) 2.37% และได้รับการทำนายผิดเป็นคลาสขั้วกลาง (Neutral Class) 8.75%

4.2.4 สถิติทดสอบที (*T*-test Statistic)

ผลลัพธ์จากตารางที่ 4.3 จะถูกนำมาทดสอบ โดยใช้สถิติทดสอบทีแบบสองกลุ่มตัวอย่างที่ไม่ขึ้นต่อกัน (Independent Two-sample *T*-test) ซึ่งแสดงให้เห็นได้ดัง ตารางที่ 4.7 โดยจากตาราง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังกล่าวนั้น จะแสดงให้เห็นถึงความแตกต่างระหว่างผลการทดลอง 2 ผลการทดลอง ที่ใช้เคอร์เนลที่แตกต่างกัน ว่าเป็นความแตกต่างที่มีนัยสำคัญ (Significant Difference) หรือไม่ ซึ่งจากตารางสามารถที่จะกล่าวได้ว่า ความแตกต่างระหว่างการทดลองที่ใช้เคอร์เนลที่แตกต่างกันนั้นมีนัยสำคัญนั้น ซึ่งสังเกตได้ที่กรณีที่ค่าพี (p) มีค่าน้อยกว่า 0.001 ยกเว้นในกรณีการเปรียบเทียบระหว่างเคอร์เนล Linear กับเคอร์เนล Polynomial ที่ใช้จำนวนคำที่นำมาดึงเป็นลักษณะสำคัญ จำนวน 3 คำ นั้นยังไม่ถือว่าเป็นความแตกต่างที่มีนัยสำคัญอย่างเด็ดขาด เนื่องจากค่าพี (p) มีค่าสูงถึง 0.7599 ด้วยกัน

ตารางที่ 4.7 ผลลัพธ์ของสถิติทดสอบที

จำนวนคำที่นำมาดึงลักษณะสำคัญ	เคอร์เนลที่ 1	เคอร์เนลที่ 2	สถิติที	ค่าพี (p)
1	Linear	RBF	-84.98	< 0.001
1	Linear	Polynomial	-37.51	< 0.001
1	RBF	Polynomial	101.65	< 0.001
2	Linear	RBF	-157.83	< 0.001
2	Linear	Polynomial	-8.92	< 0.001
2	RBF	Polynomial	141.69	< 0.001
3	Linear	RBF	-119.18	< 0.001
3	Linear	Polynomial	-0.31	0.7599
3	RBF	Polynomial	117.34	< 0.001
4	Linear	RBF	-102.83	< 0.001
4	Linear	Polynomial	28.28	< 0.001
4	RBF	Polynomial	128.65	< 0.001
5	Linear	RBF	-85.72	< 0.001
5	Linear	Polynomial	41.14	< 0.001
5	RBF	Polynomial	122.70	< 0.001
6	Linear	RBF	-72.18	< 0.001
6	Linear	Polynomial	48.84	< 0.001
6	RBF	Polynomial	117.91	< 0.001

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
การวิเคราะห์ข้อผิดพลาด

ค่าความแม่นยำเฉลี่ยสูงสุด จากการทำนายผลโดยแบ่งกลุ่มของข้อมูลประโยคในนิทาน เป็นกลุ่มสำหรับการฝึกสอน (Training Set) และกลุ่มสำหรับการทดสอบ (Testing Set) เป็นจำนวน 1,000 ครั้ง อยู่ที่ 72.58 ± 1.08% ซึ่งหมายความว่า ยังมีการทำนายผลผิดอยู่ถึง 27.42 ± 1.08% ด้วยกัน โดยในบทนี้ ผู้วิจัยจะทำการวิเคราะห์เหตุผล อันเป็นปัจจัยที่เป็นไปได้ ที่ทำให้เครื่องทำนายผลผิด โดยจะกล่าวแยกเป็นปัจจัย ๆ และให้รายละเอียดดังนี้

5.1 ข้อผิดพลาดจากการกำกับอารมณ์ให้กับประโยคในนิทานโดยใช้ผู้เชี่ยวชาญ

ประโยคในนิทานบางประโยค ยังมีความคลุมเครือ สำหรับผู้เชี่ยวชาญที่ทำการอ่าน ในการจะลงความเห็นเด็ดขาดให้กับประโยคเหล่านั้น ว่าแสดงอารมณ์ใดออกมากันแน่ ซึ่งความคลุมเครือนั้นอาจจะส่งผลให้เกิดความผิดพลาดขึ้นได้ ในบางกรณี ยกตัวอย่างประโยคที่คลุมเครือ เช่น ในนิทานหมายเลข 1 เรื่อง หนูนากับหนูบ้าน ประโยคที่ 74 มีความว่า “ยินดีต้อนรับสู่บ้านอันต่ำต้อยของฉัน หนูบ้านพูด” ซึ่งได้รับการกำหนดอารมณ์โดยผู้ทำการอ่าน ว่าเป็น 1 หรือแสดงอารมณ์ทางด้านบวกออกมา ในขณะที่ประโยคดังกล่าว มีคำว่า “ต่ำต้อย” ซึ่งเป็นคำที่แสดงอารมณ์ทางด้านลบ ออกมาอย่างเป็นอื่นไม่ได้ แต่สาเหตุที่ผู้ทำการอ่านได้เลือกว่าประโยคดังกล่าว แสดงอารมณ์ทางด้านบวกเนื่องด้วยผู้ทำการอ่านนั้น ได้รับความรู้สึกในเชิงบวก จากคำว่า “ยินดี” หรือ “ยินดีต้อนรับ” ที่อยู่ส่วนต้นของประโยค ในกรณีเช่นนี้ ผู้อ่านจึงรู้สึกคลุมเครือ จนต้องลงความเห็นอย่างใดอย่างหนึ่งไป จากสองสิ่งที่ตนรู้สึกคลุมเครือต่อมันอยู่ ซึ่งถ้าหากลงความเห็น ไม่ตรงกับสิ่งที่เครื่องได้ทำนายประโยคดังกล่าวก็จะถูกทำนายผิด

ความคลุมเครือ ณ จุดนี้ ประกอบกับการที่มีมนุษย์ผู้เชี่ยวชาญ ที่ทำการกำหนดอารมณ์ให้กับประโยคเพียงคนเดียว อาจส่งผลให้เกิดอคติ (Bias) ขึ้น อย่างหลีกเลี่ยงไม่ได้ ซึ่งอคติตรงนี้ ย่อมส่งผลต่อภาพรวมของการวิเคราะห์อารมณ์ ให้เกิดความผิดพลาดขึ้น โดยอคติ ณ ตรงนี้ เป็นสิ่งที่เกิดขึ้นจริง หลังจากได้ทำการวิเคราะห์การกำหนดค่าอารมณ์ของมนุษย์ผู้เชี่ยวชาญอีกสองคน นำมาเทียบกับมนุษย์ผู้เชี่ยวชาญคนแรก โดยการเปรียบเทียบกันระหว่างการกำหนดค่าอารมณ์ของมนุษย์ทั้งสามคน ก็พบว่า จากประโยคทั้ง 1964 ประโยคนั้น มีประโยคจำนวน 1114 ประโยค ที่ทุกคนเห็นพ้องต้องกัน ในการกำหนดค่าอารมณ์ ซึ่งคิดเป็น 56.72% ซึ่งหมายความว่ามีประโยคถึง 43.28% ที่มนุษย์เห็นต่างกัน ในการกำหนดค่าอารมณ์ และปรากฏว่ามีประโยคจำนวน 49 ประโยค หรือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.49% ที่มนุษย์เห็นต่างกันอย่างสมบูรณ์แบบ (คือคนหนึ่งเห็นว่าเป็นอารมณ์ขั้วลบ อีกคนขั้วบวก ในขณะที่อีกคนมองว่าไม่มีขั้วอารมณ์)

และดังได้กล่าวไว้ในหัวข้อ 3.2.3 ว่าในระหว่างที่มนุษย์ได้ทำการกำหนดอารมณ์ให้กับประโยค มนุษย์ก็จะทำการเลือก คำที่แสดงอารมณ์มากที่สุดตามที่มนุษย์ได้ลงความเห็น และบันทึกคำดังกล่าวไว้ ปรากฏว่า มีอยู่ถึง 124 กรณีจาก 1024 กรณี ที่คำที่มนุษย์เลือกออกมา และมองว่าเป็นคำที่ส่งผลต่ออารมณ์ของประโยคในภาพรวม กับค่าของอารมณ์ที่มนุษย์ได้ทำการกำหนดให้กับประโยคนั้น เป็นค่าที่อยู่คนละขั้วกัน ซึ่งข้อมูลดังกล่าวนี้ ทำให้เราสันนิษฐานได้ว่า เป็นสิ่งที่ทำให้เกิดข้อผิดพลาดขึ้น

5.2 ข้อผิดพลาดที่เกิดขึ้นจากการสร้างฐานข้อมูลทรัพยากรด้านอารมณ์

ฐานข้อมูลทรัพยากรด้านอารมณ์นั้น ยังมีมโนทัศน์ไม่ครอบคลุมเพียงพอ ซึ่งสามารถวิเคราะห์ได้จาก คำศัพท์เด่นที่ผู้ทำการกำกับอารมณ์ให้กับนิทานนั้น ได้เลือกขึ้นมาจากประโยค 1,024 ประโยค ประโยคละหนึ่งคำ ซึ่งสามารถนับจำนวนคำที่ผู้ใช้เลือก โดยขจัดคำที่ซ้ำซ้อนออกได้ว่า มีจำนวน 357 คำด้วยกัน (ดูภาคผนวก) และจากประโยค 1,024 เราจะสามารถพบได้ว่า มีประโยคถึง 70 ประโยค ที่ผู้ใช้ได้เลือกคำแสดงอารมณ์ออกมา แต่คำแสดงอารมณ์ในประโยคเหล่านั้น กลับไม่มีค่าอารมณ์ใด ๆ กำกับอยู่ หรือกล่าวก็คือ คำเหล่านั้น ไม่ได้อยู่ในฐานข้อมูลที่เก็บทรัพยากรทางด้านอารมณ์ในภาษาไทยไว้ ซึ่งนี่เป็นสิ่งที่ส่งผลโดยตรง ในขั้นตอนการดึงลักษณะสำคัญของประโยคออกมา เพราะจะทำให้เครื่องนั้น เลือกคำอื่น ๆ ออกมาเพื่อดึงลักษณะสำคัญ แทนที่จะเป็นคำเดียวกับที่มนุษย์ได้เลือกไว้ และส่งผลให้เกิดความผิดพลาดจากการทำนายขึ้น เช่น ในนิทานหมายเลข 1 เรื่อง หนูนากับหนูบ้าน ประโยคที่ 23 มีความว่า “เจ้าหนูจึงกระโดดลง มันกระโดดลงไปตามขั้นบันไดที่สถานีแล้วมายืนอยู่บนถนนนอกเมืองสายเล็ก ๆ ที่แสนเงียบเหงา” ซึ่งได้รับการกำหนดอารมณ์โดยผู้ทำการอ่าน ว่าเป็น -1 หรือแสดงอารมณ์ทางด้านลบ ออกมา และผู้ใช้ได้เลือกคำว่า “เงียบเหงา” เป็นคำเด่นแสดงอารมณ์ แต่ทว่าคำว่า “เงียบเหงา” นั้น กลับไม่ใช่คำที่อยู่ในฐานทรัพยากรทางด้านอารมณ์ภาษาไทยที่ได้จัดสร้างขึ้น เป็นต้น

อีกเหตุแห่งข้อผิดพลาด ที่เกิดขึ้นได้จากการสร้างฐานข้อมูลทรัพยากรด้านอารมณ์นั้น ก็คือวิธีการคิดค่าอารมณ์สำหรับแต่ละคำที่เป็นผลลัพธ์จากการแปลแบบสองทาง ที่ใช้วิธีการเฉลี่ยค่าอารมณ์แต่ละด้านของคำศัพท์ต้นทาง มาเป็นค่าอารมณ์ของคำศัพท์ปลายทาง ซึ่งนี่ไม่สมควรจะเป็นวิธีการที่ดีที่สุด เนื่องเพราะในความเป็นจริงนั้น ถึงแม้พจนานุกรมจะเก็บความหมายของคำศัพท์หนึ่ง ๆ ไว้ และคำศัพท์นั้นจะสามารถแปลกลับมาแบบสองทางได้ก็ตาม แต่คำศัพท์จากภาษาต้นทาง และปลายทาง มีความเป็นไปได้สูง ที่จะส่งผลต่อความรู้สึกต่างกัน เช่นคำว่า “Hatred” กับคำว่า “Rage” นั้น ให้คำแปลแบบสองทางเป็นภาษาไทยคือคำว่า “โทสะ” ได้เช่นเดียวกัน หากแต่ว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

"Rage" นั้นแสดงออกถึงความรู้สึกในทางโกรธ หรือไม่พอใจ ที่มากกว่าคำว่า "Hatred" และน่าจะตรงกับความหมายในภาษาไทยของคำว่า "โทสะ" มากกว่า ดังนั้นการนำคำเหล่านี้มาเฉลี่ย เพื่อกำหนดค่าอารมณ์ให้กับคำว่า "โทสะ" จึงเป็นวิธีการที่ก่อให้เกิดข้อผิดพลาดขึ้น

5.3 ข้อผิดพลาดที่เกิดขึ้นจากการดึงลักษณะสำคัญของประโยคที่ไม่มีค่าอารมณ์

ประโยคที่ไม่มีค่าอารมณ์ ในนิทานหลายประโยค สามารถสังเกตได้ว่า ประโยคเหล่านั้น กลับได้รับการเลือกคำเด่นแสดงอารมณ์ออกมา โดยเครื่อง อันเป็นคำที่มีค่าขั้วของอารมณ์ (Polarity) สูงในระดับหนึ่ง ซึ่งส่งผลต่อความผิดพลาดได้ เช่น ในนิทานหมายเลข 1 เรื่อง หนูนากับหนูบ้าน ประโยคที่ 89 มีความว่า "หลังจากที่กินเค้กช็อกโกแลตเข้าไป และเลียน้ำตาลบนชิ้นเชอรี่พร้อมกับเคี้ยวทาร์ตสตรอเบอรี่ชิ้นใหญ่อย่างติดหนึบแล้ว หนูนาก็คิดว่าบ้านในเมืองหลังนี้ให้ความรู้สึกเหมือนเป็นบ้านตัวเองเลย" ซึ่งได้รับการกำหนดอารมณ์โดยผู้ทำการอ่าน ว่าเป็น 0 หรือไม่แสดงอารมณ์ใดออกมา ออกมา ทว่าเครื่องกลับใช้ความพยายามมากเกิน จนเลือกคำว่า "เหมือน" กับ "ใหญ่" ออกมาเป็นคำเด่นแสดงอารมณ์ ซึ่งทั้งสองคำนั้นมีเซ็นติกเวกเตอร์คือ [0.13, 0.25, -0.13, 0.31, 0.19] และ [0.13, 0.19, -0.03, 0.26, 0.17] ตามลำดับ ในขณะที่ทั้งสองคำนั้น มีค่าขั้วของอารมณ์ (Polarity) อยู่ในระดับหนึ่ง อาจส่งผลให้เครื่องนั้น มองว่าประโยคดังกล่าว มีอารมณ์เป็นบวกได้

5.4 ข้อผิดพลาดที่เกิดขึ้นจากข้อจำกัดของเครื่องมือทางด้านการประมวลผลภาษาธรรมชาติ

เครื่องมือต่าง ๆ ทางด้านการประมวลผลภาษาธรรมชาติ สำหรับภาษาไทยนั้น ยังมีข้อจำกัดและต้องการการพัฒนาประสิทธิภาพ ซึ่งข้อจำกัดนี้ส่งผลให้ความผิดพลาดนั้น สามารถเกิดขึ้นได้ในขั้นตอนการเตรียมข้อมูล ตัวอย่างเช่นภาพที่ 3.2 ในส่วนผลลัพธ์ของการตัดคำ เราจะสามารถเห็นได้ว่าคำว่า "ต้นอ้อ" ถูกตัดแยกออกจากกันเป็น "ต้น|อ้อ" ทั้งที่เป็นคำเดียวกัน หรือในส่วนผลลัพธ์ของการกำหนดประเภทของคำ เราก็จะสามารถเห็นได้ว่า คำว่า "ช่าง" นั้น ถูกกำหนดประเภทของคำเป็น *ncn* ซึ่งคำว่า "ช่าง" ในวลีที่ว่า "ช่างสง่างามยิ่งนัก" นั้น ไม่ใช่คำนาม แต่ควรที่จะเป็น *adv* ที่ขยายคำว่า "สง่างาม" เช่นเดียวกันคำว่า "สง่างาม" ถูกตัดแยกออกจากกันเป็น "สง่า|งาม" ทั้งที่เป็นคำเดียวกัน ส่งผลให้คำว่า "สง่า" ถูกกำหนดอารมณ์เป็น *npn* ไป ซึ่งเป็นข้อผิดพลาดที่เกิดขึ้นต่อจากข้อผิดพลาดแรกในขั้นตอนการตัดคำนั่นเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6
สรุปผลการวิจัย และข้อเสนอแนะ

ในบทนี้ ผู้วิจัยจะทำการสรุปผลการวิจัย และเรียบเรียงสิ่งที่เป็นข้อเสนอแนะ แนวทาง ในการวิจัย และพัฒนางานทางด้านนี้ต่อ ๆ ไป

6.1 สรุปผลการวิจัย

การวิจัยในวิทยานิพันธ์ฉบับนี้ มุ่งเน้นในการจัดสร้างทรัพยากรทางด้านอารมณ์ในภาษาไทย และการวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทย โดยการจัดสร้างทรัพยากรทางด้านอารมณ์ในภาษาไทยนั้น ได้ใช้วิธีการแปลแบบสองทาง ในการแปลมโนทัศน์ตั้งต้น ที่นำมาจากเซ็นติกเน็ต 2 (SenticNet 2) มาเป็นฐานข้อมูลทรัพยากรทางอารมณ์ภาษาไทย ซึ่งจากมโมทัศน์ตั้งต้นจำนวน 14,244 มโนทัศน์นั้น ได้มโนทัศน์ภาษาไทยเป็นจำนวน 16,478 มโนทัศน์ด้วยกัน และสำหรับการวิเคราะห์อารมณ์ในนิทานเด็กภาษาไทยนั้น ได้มีการนำเสนอเวกเตอร์ของลักษณะสำคัญที่ใช้กับซัพพอร์ตเวกเตอร์แมชชีพ ซึ่งความแม่นยำสูงสุดเฉลี่ยจากการทดลองจำนวน 1,000 ครั้งนั้น ได้ผลลัพธ์ที่ 72.58 ± 1.08% โดยการใช้ เคอร์เนล RBF กับค่าพารามิเตอร์ $C = 1.0$ และ $Gamma = 0.1$

6.2 ข้อเสนอแนะ

จากบทที่ 5 นั้น จะเห็นได้ว่า ระบบยังมีความต้องการในการพัฒนาเพิ่มเติมให้มากขึ้นในส่วนต่าง ๆ แทบจะทุกส่วน ทุกขั้นตอน เพื่อขจัดข้อผิดพลาดที่เกิดขึ้นที่ส่งผลถึงผลลัพธ์ท้ายสุด และเพื่อที่จะได้รับแบบจำลองการเรียนรู้ของเครื่องที่เหมาะสม มีประสิทธิภาพ และดีที่สุด ในการนำมาทำนายอารมณ์จากประโยค ในนิทานสำหรับเด็ก ซึ่งข้อเสนอแนะ สำหรับการต่อยอดในงานวิจัยด้านนี้ต่อ ๆ ไปนั้น มีดังนี้

- ทรัพยากรด้านอารมณ์ในภาษาไทยนั้น ยังสามารถที่จะพัฒนาเพิ่มเติมได้อีก ด้วยหนทางต่าง ๆ เช่น การนำพจนานุกรมฉบับอื่น ๆ มาร่วมในการแปลมโนทัศน์จากภาษาอังกฤษ สู่ภาษาไทย หรือ การแปลโดยใช้ตัวแปลของกูเกิ้ล (Google Translate) เป็นต้น
- การรวบรวมคำขยายต่าง ๆ (Modifier) หรือคำที่ส่งผลให้เกิดความหมายทางด้านตรงกันข้าม หรือคำเชิงปฏิเสธ (Negation) ต่าง ๆ อันเป็นกลุ่มคำที่มีผลให้ค่าอารมณ์ของประโยคแปรเปลี่ยน หรือกลับด้าน เพื่อที่จะนำคำเหล่านั้น มามีส่วนร่วมในขั้นตอนการดึงลักษณะเด่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ในขั้นตอนการดึงลักษณะสำคัญ ที่มีจุดบกพร่องในกรณีของประโยคที่มีค่าอารมณ์เป็นกลาง แต่กลับมีคำศัพท์ที่แสดงออกถึงอารมณ์ออกมา อันส่งผลให้การเรียนรู้ของเครื่อง สามารถมองประโยคนั้น ๆ ว่า เป็นประโยคที่มีขั้วของอารมณ์ได้ ซึ่งก่อให้เกิดการสร้างแบบจำลองที่ผิด และการทำนายผลที่ไม่ถูกต้อง ซึ่งเป็นจุดยากท้าทายจุดหนึ่งที่จะต้องหาทางแก้ไขปัญหาให้ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

Bird, S., Ewan, K., & Loper, E. (2009). *Natural Language Processing with Python.* Sebastopol: O'Reilly Media.

Cambria, E., & Hussain, A. (2012). *Sentic Computing: Techniques, Tools, and Applications.* Springer.

Cambria, E., Havasi, C., & Hussain, A. (2012). SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis. *FLAIRS conference*, (pp. 202-207). florida.

Cambria, E., Livingstone, A., & Hussain, A. (2012). The hourglass of emotions. *Cognitive behavioural systems* (pp. 144-157). Berlin Heidelberg: Springer.

Charoenpornsawat, P. (2003). *SWATH: Smart Word Analysis for Thai.* Retrieved from Carnegie Mellon University: http://www.cs.cmu.edu/~paisarn/software.html

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research 16*, 321–357.

Cournapeau, D., & Brucher, M. (2010). Retrieved from Scikit-learn machine learning in Python: http://scikit-learn.org/

Haruechaiyasak, C., Kongthon, A., Pornpimon, P., & Kanokorn, T. (2013). S-Sense: A Sentiment Analysis Framework for Social Media Sensing. *Sixth International Joint Conference on Natural Language Processing*, (p. 6). Nagoya.

Jitar model and Jitar. (2010, 24 2). Retrieved from NAiST Lab, Kasetsart University: http://naist.cpe.ku.ac.th/pkg/

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *The International Joint Conference on Artificial Intelligence (IJCAI)*, *14*, pp. 1137-1145.

Lertsuksakda, R., Netisopakul, P., & Pasupa, K. (2014). Thai Sentiment Terms Construction using the Hourglass of Emotions. *the 6th International Conference on Knowledge and Smart Technology (KST 2014)*, (pp. 46-50). Chonburi.

Lertsuksakda, R., Pasupa, K., & Netisopakul, P. (2015). Sentiment Analysis on Thai Children Stories with Support Vector Machine. *the 20th International Symposium on Artificial Life and Robotics (AROB 2015)*, (pp. 138-142). Beppu.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing.* london: The MIT press.

Mudinas, A., Zhang, D., & Levene, M. (2012). Combining lexicon and learning based approaches for concept-level. *the 1st International Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM 2012)*, (pp. 1-8). Beijing.

Nathalie, N. (2000). The class imbalance problem: Significance and strategies. *Proc. of the Int'l Conf. on Artificial Intelligence.*

Netisopakul, P., & Wikaha, P. (2014). Speaker identification from multiple short stories without an avatar list. *The 10th Symposium on Natural Language*, (pp. 1-7). Phuket.

Plutchik, R. (1987). The nature of emotions. In A. Ben-Zeev, *Philosophical Studies* (pp. 393-409).

Satayamas, V. (2009, 11 23). *VEER66.* Retrieved from KUCut API: https://veer66.wordpress.com/2009/11/23/kucut_api/

Vapnik, V., & Cortes, C. (1995). Support-vector networks. *Machine learning, 20*, 273-297.

Wang, J., & Oard, D. W. (2006). Combining bidirectional translation and synonymy for cross-language information retrieval. *the 29th Annual SIGIR Conference*, (pp. 202-209). Seattle.

ชูเดช, ส. (2008). *เอกสารจิตวิทยาสำหรับนักศึกษา/เอกสารประกอบการสอนบทที่ 12.* Retrieved from มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี: arts.kmutt.ac.th/ssc231/document/เอกสารจิตวิทยาสำหรับนักศึกษา/เอกสารประกอบการสอนบทที่ 12.doc

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก.

ตัวอย่างทรัพยากรทางด้านอารมณ์ภาษาไทย

ตัวอย่างทรัพยากรทางด้านอารมณ์ภาษาไทยจำนวน 357 คำ หลังจากกำจัดคำที่ซ้ำกันออกไป ซึ่งเป็นคำที่มนุษย์มองว่าเป็นคำเด่นแสดงอารมณ์ของแต่ละประโยค และได้เลือกมันขึ้นมา จากประโยคทั้งหมดจำนวน 1,024 ประโยค ซึ่งเป็นกลุ่มประโยคที่แสดงขั้วของอารมณ์ออกมาเป็นบวก (Positive Class) หรือเป็นลบ (Negative Class)

คำศัพท์	ค่าความพึงพอใจ (Pleasantness)	ค่าความสนใจ (Attention)	ค่าความอ่อนไหว (Sensitivity)	ค่าการยอมรับ (Aptitude)	ค่าขั้วของอารมณ์ (Polarity)
สุข	0.76	0.03	-0.03	0.16	0.31
วุ่นวาย	-0.42	0.3	-0.27	-0.12	-0.17
วิเศษ	0.41	-0.07	-0.07	0.41	0.41
ร้อน	0.11	0.2	-0.13	0.19	0.12
จริง	0.18	0.35	-0.09	0.36	0.27
ยินดี	0.77	0.47	-0.14	0.38	0.49
เงียบเหงา	0	0	0	0	0
ดีใจ	0.72	0.44	-0.2	0.24	0.4
ผิดหวัง	-0.65	0.25	-0.4	-0.41	-0.4
ดี	0.48	0.55	-0.02	0.27	0.36
สบาย	0.8	0.18	0	0.18	0.38
ตื่นเต้น	0.24	0.82	-0.51	0.12	0.22
แปลกใจ	0.21	-0.4	0	0.33	0.43
จ้ะ	0.47	0.55	-0.07	0.57	0.51
ประหลาด	0.05	-0.44	-0.01	0.23	0.26
ตกใจ	-0.47	0.35	-0.22	-0.18	-0.17
ที่รัก	0.29	0.25	-0.26	0.37	0.22
หนักแน่น	0.4	0.08	0	0.58	0.35
กลัว	0	0	-0.89	-0.4	-0.43
สั่น	0.01	-0.01	-0.39	0.05	-0.1
เหน็ดเหนื่อย	0	0	0	0	0
งดงาม	0.58	0.34	-0.03	0.46	0.45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำศัพท์	ค่าความพึงพอใจ (Pleasantness)	ค่าความสนใจ (Attention)	ค่าความอ่อนไหว (Sensitivity)	ค่าการยอมรับ (Aptitude)	ค่าขั้วของอารมณ์ (Polarity)
หัวเราะ	0.63	0.42	-0.03	0.21	0.41
โอ้อวด	0.32	0	0.32	-0.81	-0.27
อร่อย	0.3	0.3	-0.02	0.33	0.3
ร่าเริง	0.76	0.52	-0.14	0.38	0.51
หอม	0.49	0.46	0	0.09	0.34
ปลอดภัย	0.02	0.31	-0.1	0.21	0.15
ดุร้าย	0	0	0	0.08	0.03
อลหม่าน	-0.56	0.38	-0.28	-0.14	-0.2
โกรธ	0	0	0.86	0	-0.29
เจ็บปวด	-0.59	-0.14	-0.44	-0.16	-0.32
เสีย	-0.11	0.09	-0.06	-0.01	-0.02
เสี่ยงชีวิต	0	0	0	0	0
อุ่น	0.22	0.28	-0.14	0.12	0.16
น่ากลัว	-0.32	0.05	-0.43	-0.22	-0.28
สิ้นหวัง	0	0	0	0	0
เจ็บ	-0.61	0	-0.13	-0.51	-0.6
รบกวน	-0.25	0.05	0.27	-0.28	-0.28
น้ำใจ	0.21	0.21	-0.16	0.41	0.23
ร้อง	-0.09	0.1	0	-0.15	-0.08
โง่	-0.16	0	0.44	-0.6	-0.4
สะอาด	0.3	0.05	-0.03	0.45	0.26
แปลก	0.05	-0.22	-0.02	0.1	0.17
คำราม	0	0	-0.87	0	-0.29
ลำบาก	-0.16	0.06	-0.15	0.07	-0.06
ช่วย	0.49	0.38	-0.02	0.66	0.5
ตลก	0.2	0.03	0.06	0.04	0.05
ยิ้ม	0.57	-0.01	0.01	-0.02	0.18
เดือดร้อน	-0.52	0.41	-0.01	-0.42	-0.27
สงสาร	0	0	0	0.82	0.27
หวัง	0.32	0.75	-0.12	0.11	0.35
เยาะ	-0.22	0.03	0.24	0	-0.25
อิสระ	0.88	0.81	0	0.89	0.86

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำศัพท์	ค่าความพึงพอใจ (Pleasantness)	ค่าความสนใจ (Attention)	ค่าความอ่อนไหว (Sensitivity)	ค่าการยอมรับ (Aptitude)	ค่าขั้วของอารมณ์ (Polarity)
แข็งแรง	0.31	0.18	-0.06	0.29	0.24
กระตือรือร้น	0.4	0.54	-0.34	0.08	0.23
สง่า	0.04	0.51	0.01	0.53	0.36
ขอบคุณ	0.86	0	0	0.96	0.61
ดูถูก	-0.52	0.09	0.77	-0.76	-0.65
เจ้าเล่ห์	-0.38	0.17	-0.32	0.24	-0.1
กิตติศัพท์	0.66	0.5	-0.02	0.7	0.61
ฉลาด	0.23	0.29	-0.02	0.03	0.18
สุภาพ	0.03	0.05	-0.05	0.53	0.19
พักผ่อน	0.45	0.34	0.01	0.44	0.43
ต้อนรับ	0.44	0.25	-0.07	0.47	0.37
เศร้า	-0.85	0.18	-0.44	-0.01	-0.37
ยิ้มแป้น	0.71	0	0	0.77	0.49
หลอก	-0.2	-0.01	-0.13	-0.19	-0.19
เอาชนะ	0.03	0.25	-0.21	0.27	0.18
แน่ใจ	0.18	0.22	-0.09	0.64	0.32
โมโห	-0.05	0.05	0.78	-0.5	-0.43
อันตราย	-0.25	-0.02	-0.26	-0.33	-0.32
พึมพำ	0	0.06	0.49	-0.31	-0.25
สำคัญ	-0.03	0.22	-0.1	0.2	0.1
พระทัย	0.1	0.29	-0.17	0.14	0.12
หนัก	-0.05	0.22	-0.09	0.16	0.08
ช่วยเหลือ	0.5	0.36	-0.09	0.57	0.45
ชื่นชอบ	0.47	0.33	-0.2	0.37	0.32
ป้องกัน	0.01	0.27	-0.19	0.21	0.1
สัญญา	0.18	0.22	-0.17	0.44	0.22
เห็นชอบ	0	0	0	0	0
เรียบร้อย	0.57	0.07	-0.04	0.41	0.33
ประณีต	0.46	0.27	-0.07	0.47	0.38
ต่อย	0	0.04	-0.05	-0.05	0
เอาเป็นเอาตาย	0.06	0.62	-0.31	0.14	0.17
โอ๊ย	0	0	0	0	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำศัพท์	ค่าความพึงพอใจ (Pleasantness)	ค่าความสนใจ (Attention)	ค่าความอ่อนไหว (Sensitivity)	ค่าการยอมรับ (Aptitude)	ค่าขั้วของอารมณ์ (Polarity)
สำเร็จ	0.48	0.41	-0.05	0.54	0.46
ลำพอง	0	0	0	0	0
หน้าซีด	0	0	0	0	0
แย่	-0.71	-0.07	-0.66	-0.34	-0.52
ร้องไห้	-0.45	-0.04	-0.08	-0.55	-0.35
คร่ำครวญ	-0.8	0	0.38	-0.37	-0.52
ตะโกน	-0.08	-0.02	-0.48	-0.02	-0.19
อาย	-0.5	0.39	-0.22	-0.16	-0.16
เสียชีวิต	-0.15	0.09	-0.21	-0.08	-0.12
ทุกข์	-0.52	0.17	-0.23	-0.2	-0.31
เล่น	0	0	0	0	0
เก่ง	0.23	0.1	-0.08	0.09	0.11
ฝัน	0.36	0.52	-0.03	0.38	0.41
คราง	-0.06	-0.01	-0.47	-0.5	-0.34
ลืม	-0.88	0	0.46	-0.49	-0.61
สวย	0.35	0.4	-0.02	0.4	0.38
ปัญหา	-0.41	-0.2	-0.28	-0.29	-0.24
โอกาส	0.21	0.3	-0.11	0.18	0.19
กล้าหาญ	0.11	0.18	-0.09	0.53	0.24
ระมัดระวัง	0.1	0.2	-0.12	0.17	0.12
มุ่งมั่น	0	0	0	0	0
เกียจคร้าน	0	0	0.9	-0.76	-0.55
ตาย	-0.17	0.11	-0.12	-0.07	-0.09
อกสั่นขวัญแขวน	0	0	0	0	0
พอใจ	0.77	0.38	-0.15	0.19	0.4
สะอึกสะอื้น	-0.71	0.03	-0.33	-0.45	-0.49
ชีวิต	0.53	0.48	-0.1	0.57	0.5
ปาฏิหาริย์	0.82	-0.84	0	0.7	0.78
อ่อนแอ	-0.11	0.03	-0.12	0.11	-0.01
เบิกบาน	0.68	0.66	-0.21	0.35	0.49
มั่นคง	0.04	0.12	-0.01	0.1	0.08
นอบน้อม	0.32	0.46	-0.35	0.44	0.29

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำศัพท์	ค่าความพึงพอใจ (Pleasantness)	ค่าความสนใจ (Attention)	ค่าความอ่อนไหว (Sensitivity)	ค่าการยอมรับ (Aptitude)	ค่าขั้วของอารมณ์ (Polarity)
ดื้อ	-0.02	0.11	0.24	-0.22	-0.16
ซื่อสัตย์	0.47	0.43	0.14	0.6	0.44
ตั้งอกตั้งใจ	0	0	0	0	0
อ่อนโยน	0.07	0.24	-0.08	0.4	0.21
เมตตา	0.53	0.59	-0.07	0.7	0.58
อดตาย	-0.23	0.05	-0.28	-0.03	-0.17
โชคดี	0.99	0.86	0	0.9	0.92
โชค	0.97	0.41	0.39	0.44	0.48
ใช่	0.47	0.55	-0.07	0.57	0.51
โลภ	0.07	0.06	0.83	-0.73	-0.47
ใส่ใจ	0	0.12	-0.06	0.2	0.09
เงียบ	0.34	0.38	-0.18	0.46	0.33
ตามลำพัง	-0.81	0	-0.9	0	-0.57
ทำใจ	0.25	0.2	-0.09	0.35	0.24
น้ำตา	-0.11	-0.04	-0.13	-0.15	-0.12
สวยงาม	0.39	0.31	-0.06	0.36	0.33
ปลื้มใจ	0.89	0.97	0	0.91	0.92
ประหยัด	-0.03	0.34	-0.11	0.25	0.15
งาม	0.29	0.35	-0.07	0.38	0.32
ใจดี	0.6	0.41	-0.25	0.52	0.43
รัก	0.38	0.16	-0.19	0.38	0.24
สนุกสนาน	0.76	0.51	-0.12	0.16	0.44
ครับ	0.47	0.55	-0.07	0.57	0.51
บันดาลใจ	0	0.96	0	0.96	0.64
สนุก	0.79	0.57	0	0.33	0.56
ซาบซึ้ง	0.28	0.14	-0.1	0.3	0.25
ชอบ	0.41	0.51	-0.1	0.42	0.41
เอร็ดอร่อย	0.08	0.03	0.01	0.13	0.07
สดชื่น	0.39	0.37	-0.16	0.21	0.27
ชื่นชม	0.5	0.14	0.2	0.8	0.42
ความสำเร็จ	0.4	0.1	-0.02	0.15	0.21
ไชโย	0	0	0	0	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำศัพท์	ค่าความพึงพอใจ (Pleasantness)	ค่าความสนใจ (Attention)	ค่าความอ่อนไหว (Sensitivity)	ค่าการยอมรับ (Aptitude)	ค่าขั้วของอารมณ์ (Polarity)
เกเร	0	0	0.93	-0.92	-0.62
แกล้ง	-0.24	-0.02	0.48	-0.47	-0.39
จรจัด	-0.89	0	0	-0.74	-0.54
กลั่นแกล้ง	0	0	0	0	0
ตก	-0.35	-0.15	-0.12	-0.3	-0.28
สบายใจ	0.82	0.25	0	0	0.36
เพื่อน	0.3	0.1	-0.05	0.25	0.2
ล้อ	-0.03	0.19	0.18	-0.11	-0.03
ล้อเลียน	-0.18	0.12	-0.2	0.16	-0.03
เคร่งขรึม	-0.11	0.13	-0.12	-0.11	-0.07
เงียบกริบ	0.58	0.63	-0.23	0.27	0.42
คลาน	-0.43	0.36	-0.18	-0.13	-0.13
กำลังใจ	0.18	0.1	-0.07	0.23	0.14
มั่นใจ	0.28	0.33	-0.15	0.6	0.35
แจ่มใส	0.45	0.51	-0.09	0.29	0.39
พัก	0.12	0.13	-0.03	0.16	0.13
เชียร์	0.39	0.25	0.06	0.4	0.24
ภาคภูมิใจ	0	0	0	0	0
ชัย	0.82	0.63	0	0.58	0.68
เร่ง	0.37	0.43	-0.19	0.26	0.29
ไม่ใช่	0	0	0	0	0
เหลือเชื่อ	0	0	0	0	0
ปรบมือ	0.15	-0.06	0	-0.09	0.04
กว้าง	0	0.09	0.01	0.18	0.07
ขายหน้า	-0.56	0.33	-0.32	-0.28	-0.28
ภูมิใจ	0.71	0.49	-0.24	0.37	0.44
นุ่มนวล	0.12	0.15	-0.04	0.39	0.21
เห็นด้วย	0.38	0.48	-0.11	0.41	0.38
ขมวดคิ้ว	-0.19	0	-0.24	0	-0.14
หงุดหงิด	-0.48	0.41	-0.19	-0.19	-0.15
แพ้	-0.08	-0.09	0.06	-0.04	-0.03
เลว	-0.14	0.06	0.05	-0.49	-0.35

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำศัพท์	ค่าความพึงพอใจ (Pleasantness)	ค่าความสนใจ (Attention)	ค่าความอ่อนไหว (Sensitivity)	ค่าการยอมรับ (Aptitude)	ค่าขั้วของอารมณ์ (Polarity)
เคียดแค้น	0	0	0	0	0
นิทาน	-0.02	0.48	-0.02	0	0.14
ยากจน	0	0	0	0	0
เจ็บป่วย	0	0	0	0	0
มอซอ	0	0	0.94	-0.8	-0.58
ล่า	-0.04	0.02	-0.04	-0.06	-0.05
อบอุ่น	0.5	0.63	-0.07	0.52	0.52
ทะเลาะ	-0.85	0	0.87	-0.97	-0.9
หิว	0.08	0.23	-0.25	0.2	0.09
กรีดร้อง	0	0	-0.81	0	-0.27
โหยหวน	0	0	0	0	0
บุญ	0.73	0.91	0	0.79	0.81
งุนงง	-0.04	-0.55	-0.06	0.31	0.25
ปีศาจ	-0.04	0.04	-0.41	-0.49	-0.3
แห้งแล้ง	0	-0.05	0	0.05	0.03
เวิ้งว้าง	-0.03	0.1	0.01	0.17	0.07
เสด็จ	-0.01	0.25	-0.16	0.37	0.15
โกลาหล	-0.8	-0.37	-0.87	0	-0.43
ประโยชน์	0.22	0.52	-0.07	0.36	0.34
บริสุทธิ์	0.02	0.04	-0.02	0.32	0.12
มิตร	0.74	0	0	0.44	0.39
ชุ่มชื่น	0.66	0.64	-0.24	0.3	0.45
บ้า	-0.05	0	0.48	-0.64	-0.42
อารมณ์	0	0.07	0.29	-0.3	-0.21
เถียง	-0.39	0	0.79	-0.86	-0.68
ชน	-0.05	0.01	-0.02	-0.03	-0.02
สุดกำลัง	0	0	0	0	0
โวยวาย	0	0	0	0	0
ตำหนิ	-0.23	0.03	-0.13	-0.08	-0.13
สงบ	0.36	0.36	-0.15	0.32	0.3
สติ	-0.01	0.14	-0.08	0.42	0.15
หลับตา	0	0	0	0	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำศัพท์	ค่าความพึงพอใจ (Pleasantness)	ค่าความสนใจ (Attention)	ค่าความอ่อนไหว (Sensitivity)	ค่าการยอมรับ (Aptitude)	ค่าขั้วของอารมณ์ (Polarity)
อ่อนแรง	0	0.14	0.01	0.14	0.09
ลืมตา	0	0	0	0	0
อุทาน	0	0.16	-0.1	0.3	0.12
ดุ	-0.19	0.02	-0.15	-0.15	-0.22
ขอร้อง	0	0.37	0.35	-0.49	-0.16
ป่วย	-0.57	0.03	0.11	-0.75	-0.69
ใหม่	0.04	0.15	-0.09	0.17	0.09
เคารพ	0.41	0.65	-0.18	0.56	0.48
หลง	-0.55	0.03	0.25	-0.21	-0.35
อิจฉา	0.03	0.03	0.87	-0.47	-0.43
ห่วง	-0.21	0.46	-0.29	-0.06	-0.03
ห่วงใย	0	0	0	0	0
ทิ้ง	0	0	0	0.7	0.23
ดูแล	0.15	0.19	-0.11	0.16	0.13
ขะมักเขม้น	0.35	0.36	-0.15	0.17	0.24
เชื่อใจ	0	0	0	0.72	0.24
หมดหวัง	-0.66	0.21	0.37	-0.51	-0.51
หกล้ม	-0.07	-0.52	-0.38	-0.04	0.01
ขอโทษ	-0.12	0.18	-0.18	0.23	0.04
กังวล	-0.22	0.25	-0.28	-0.09	0.02
รังเกียจ	-0.21	0.1	0.58	-0.57	-0.44
วิตก	-0.25	0.12	-0.31	-0.16	-0.12
เหงา	-0.05	0	-0.13	-0.03	-0.07
สนิทสนม	0.78	0.97	0	0.91	0.89
อาลัยอาวรณ์	-0.72	0.05	0.15	-0.52	-0.59
เบื่อ	0	0	0	0	0
เสียใจ	-0.87	0.06	-0.41	-0.11	-0.45
เต็ม	0.1	0.29	-0.11	0.24	0.17
บาดเจ็บ	-0.93	-0.99	-0.91	0	-0.29
เลี้ยง	0.31	0.1	-0.06	0.26	0.2
รักษา	0.1	0.16	-0.11	0.23	0.13
ขอบใจ	0.86	0	0	0.96	0.61

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำศัพท์	ค่าความพึงพอใจ (Pleasantness)	ค่าความสนใจ (Attention)	ค่าความอ่อนไหว (Sensitivity)	ค่าการยอมรับ (Aptitude)	ค่าขั้วของอารมณ์ (Polarity)
ของขวัญ	0.52	0.33	-0.09	0.55	0.48
พร	0	0	0	0.83	0.28
น่ารัก	0.44	0.47	-0.02	0.5	0.46
โกหก	-0.05	0.03	-0.05	0.11	0.01
ว่า	-0.09	0.04	-0.22	-0.08	-0.12
หลับ	0	0	0	0	0
เย่อหยิ่ง	-0.72	0	0.71	0	-0.48
สมควร	0	0	0	0	0
เกลียด	-0.3	0.09	0.61	-0.76	-0.53
จริงใจ	0.38	0.45	0	0.69	0.5
ผิด	-0.34	0.04	0.31	-0.45	-0.38
ประกาย	0.05	0.07	-0.06	0.12	0.05
ทน	-0.02	0.08	-0.05	0.07	0.03
เอาใจ	0.23	0.1	-0.14	0.15	0.11
ทะนง	0.71	0.49	-0.24	0.37	0.44
โชคร้าย	0	0	0	0	0
เอาแต่ใจ	0	0	0	0	0
หนักใจ	-0.33	0.28	-0.19	-0.12	-0.12
คลั่ง	-0.02	-0.01	0.36	-0.48	-0.3
หัวเราะเยาะ	-0.13	0.02	0.17	-0.08	-0.19
เข้าใจ	-0.05	0.03	-0.03	0	-0.01
เพลิดเพลิน	0.84	0.91	0	0.48	0.74
โจมตี	-0.06	-0.01	-0.07	-0.43	-0.18
เหนื่อย	-0.78	0.27	-0.46	-0.05	-0.34
คิด	-0.02	0.13	-0.01	0.16	0.06
เรี่ยวแรง	0	0.14	-0.15	0.13	0.04
ทรุด	-0.13	-0.05	0	-0.09	-0.06
ภัย	-0.38	-0.08	-0.3	-0.36	-0.45
ง่ายดาย	0.94	0	0	0	0.31
อัศจรรย์	0.3	-0.19	0	0.5	0.51
โหดร้าย	-0.5	0.37	-0.21	-0.17	-0.17
ปลอบโยน	0.65	0.34	-0.15	0.2	0.34

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำศัพท์	ค่าความพึงพอใจ (Pleasantness)	ค่าความสนใจ (Attention)	ค่าความอ่อนไหว (Sensitivity)	ค่าการยอมรับ (Aptitude)	ค่าขั้วของอารมณ์ (Polarity)
ง่าย	0.87	0	0	0	0.29
โต	0.13	0.21	-0.01	0.31	0.2
เขียว	-0.02	0.1	0.25	-0.2	-0.12
งอกงาม	-0.02	0.14	-0.03	0.14	0.07
ร่มเย็น	0.52	0.44	-0.18	0.28	0.35
ชื่อเสียง	0.71	0.56	-0.02	0.75	0.67
แผด	-0.07	-0.04	-0.48	-0.07	-0.19
ยิ่งใหญ่	0.19	0.32	-0.05	0.53	0.33
ถ่อม	0.96	0	0	0.97	0.64
ตะคอก	-0.22	0	-0.2	0	-0.14
พิเศษ	0.03	-0.07	-0.09	0.26	0.16
เชื่อ	0.08	0.34	-0.21	0.4	0.2
แพง	0.1	0.26	-0.19	0.23	0.13
ขโมย	-0.45	0	0.35	-0.41	-0.45
ซอมซ่อ	0	0	0	0	0
บันดาล	0	0.2	-0.07	0.23	0.12
ยอมรับ	0.24	0.1	-0.04	0.54	0.28
อยาก	0.35	0.44	-0.18	0.17	0.26
เสียน้ำใจ	0	0	0	0	0
ดีอกดีใจ	0	0	0	0	0
หนาว	0.23	0.24	-0.09	0.13	0.17
ร้าย	-0.3	0.01	-0.34	-0.32	-0.38
ครวญ	-0.33	0.03	0.11	-0.62	-0.56
ผอมโซ	0	0	0	0	0
วิงวอน	0.31	0.45	-0.23	0.25	0.26
ใจแคบ	0	0	0	-0.92	-0.31
ท้อแท้	-0.56	0.27	0.05	-0.38	-0.24
ชรา	0.01	0	-0.05	0.35	0.1
เข้มแข็ง	0.36	0.03	0	0	0.13
หรูหรา	0.28	0.36	0.01	0.33	0.32
สงสัย	0.05	0.25	-0.19	-0.01	0.04
ตื้นตัน	0	0	0	0	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำศัพท์	ค่าความพึงพอใจ (Pleasantness)	ค่าความสนใจ (Attention)	ค่าความอ่อนไหว (Sensitivity)	ค่าการยอมรับ (Aptitude)	ค่าขั้วของอารมณ์ (Polarity)
แต่งงาน	0.47	0.5	-0.05	0.54	0.49
บันเทิง	0.7	0.54	0	0.65	0.63
เฮ้อ	0	0	0	0	0
สุก	0	0.14	-0.07	0.22	0.1
มหัศจรรย์	0.19	-0.22	-0.04	0.44	0.39
เคร่งเครียด	-0.06	0.09	-0.12	0.02	-0.02
ช้า	-0.09	0.08	-0.08	0	-0.03
เครียด	-0.1	0.1	0.13	0.04	-0.1
ไม่ทัน	-0.95	0	0.73	-0.99	-0.89
ทำร้าย	-0.48	-0.5	-0.23	-0.48	-0.39
เชื่อฟัง	0.06	0.27	-0.21	0.34	0.15
พยายาม	-0.04	0.07	-0.09	0	-0.01
เสียดาย	-0.92	0.1	-0.72	0.1	-0.48
โหด	-0.5	0.38	-0.22	-0.16	-0.17
หล่น	-0.03	-0.07	0.07	-0.05	-0.03
ปลอบ	0.63	0.2	-0.1	0.13	0.28
ถูก	0.02	0.29	-0.2	0.32	0.14
กระวนกระวาย	-0.27	0.31	-0.08	-0.18	-0.09
จำ	-0.18	0.13	-0.36	-0.23	-0.22
เพลิน	0.7	0.95	0	0	0.55
รุนแรง	-0.1	0.01	-0.04	-0.02	-0.05
หนี	-0.03	0.02	0.02	0.08	0.03
ร่ำรวย	0.49	0.5	-0.07	0.06	0.33
ปล้น	-0.38	0.13	0.38	-0.36	-0.35
มัจจุราช	0	0	0	0	0
ชั่วร้าย	-0.38	0.1	-0.22	-0.38	-0.45
หาย	-0.08	-0.09	0.06	-0.04	-0.03
สยอง	-0.49	0.13	-0.54	-0.32	-0.41
ผลัก	-0.07	0.04	-0.07	0.07	0
โยน	-0.04	0	-0.05	0.1	0
หมด	-0.1	0.14	-0.16	0.05	-0.02
ชาญฉลาด	0.43	0.5	-0.02	0.24	0.38

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำศัพท์	ค่าความพึงพอใจ (Pleasantness)	ค่าความสนใจ (Attention)	ค่าความอ่อนไหว (Sensitivity)	ค่าการยอมรับ (Aptitude)	ค่าขั้วของอารมณ์ (Polarity)
กัด	-0.23	-0.02	0.29	-0.36	-0.31
แต่	-0.01	0.23	-0.11	0.26	0.12
ยาก	-0.05	0	-0.06	0.14	0.01
เศรษฐี	0	0	0	0	0
เหี้ยมโหด	-0.5	0.37	-0.21	-0.17	-0.17
รื่นเริง	0.7	0.68	-0.19	0.32	0.5
สรรเสริญ	0.26	0	0.33	0.74	0.22
สะดุ้ง	0	0	0	0	0
สัตว์ประหลาด	0	0	0	0	0
กล้า	0.19	0.28	-0.12	0.47	0.27
ตะลึงพรึงเพริด	0	-0.84	0	0	0.28
ดุดัน	0	0	0	0	0
พิโรธ	-0.14	0.14	0.62	-0.64	-0.42
ร้องเพลง	0.92	0	0	0	0.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข.

ตัวอย่างนิทานเรื่องหนูบ้านกับหนูนา จากนิทาน 40 เรื่อง

ประโยคที่	ข้อความในประโยค	ค่าอารมณ์
1	หนูบ้านกับหนูนา	0
2	“กินอาหารกลางวันที่ต่างจังหวัด	0
3	น่าสนใจเป็นที่สุด!”	1
4	หนูบ้านอุทานเมื่อได้รับบัตรเชิญจากญาติของมัน	0
5	ซึ่งผู้นำข่าวนี้มาแจ้งก็คือคุณไปรษณีย์นกนั่นเอง	0
6	“อากาศสดชื่น แสงแดดอ่อน ๆ มีแต่ความเงียบและสงบสุข...”	1
7	หนูน้อยยังคงฝันเฟื่องต่อไป ขณะที่มองออกไปนอกหน้าต่างแล้วเห็นความวุ่นวายตรงถนนข้างล่าง	-1
8	“เชื่อเลยว่าอาหารต้องมีรสชาติวิเศษสุด เวลาที่เราพักผ่อนอยู่ในสวนดอกไม้ใต้ท้องฟ้ากว้าง”	1
9	“ขอโทษนะจ๊ะ” นกไปรษณีย์พูด	0
10	มันกำลังคอยอยู่ที่ขอบหน้าต่างอย่างร้อนใจ	-1
11	“ต้องตอบจดหมายไหม” “โอ จริงสินะ” หนูบ้านพูด	1
12	“ช่วยบอกเขาด้วยว่าฉันยินดีที่จะไป ฉันจะไปถึงที่นั่นตอนเที่ยงวันพรุ่งนี้”	1
13	นกไปรษณีย์พยักหน้าแล้วบินหายไปในท้องฟ้า	0
14	เช้าตรู่วันรุ่งขึ้น หนูน้อยลอดตัวออกจากประตูบ้านในเมืองหลังที่มันอาศัยอยู่	0
15	แล้วออกเดินทางไปเยี่ยมญาติของมันที่ต่างจังหวัด	0
16	หนูน้อยวิ่งไปตามถนนที่มีแต่ความจอแจ มันแทรกตัวผ่านเท้าของผู้คนอย่างเร่งรีบ จนมาถึงสถานีรถไฟ	0
17	ผู้โดยสารนับร้อยเคลื่อนตัวออกจากรถไฟไปสู่ชานชาลา	0
18	พวกเขาต่างถือกระเป๋าเดินทางและร่มอยู่ในมือ	0
19	หนูน้อยหลบหลีกขาของคนเหล่านั้นอย่างรวดเร็ว แล้วกระโดดขึ้นรถไฟได้ทันเวลาพอดี	0
20	ประตูปิดตามหลังมันและรถไฟก็เคลื่อนตัวออกจากสถานี	0
21	เจ้าหนูไต่ขึ้นไปยังที่ว่างข้างหน้าต่าง มันมองภาพเมืองอันวุ่นวายที่วิ่งผ่านไป บ้านและร้านรวง ตรอกซอยและถนน ผู้คนและเสาไฟ รถประจำทางและรถยนต์ขับผ่านไปอย่างรวดเร็ว	-1
22	ภาพเหล่านั้นค่อย ๆ เปิดทางให้เนินเขาสีเขียวสูง ๆ ต่ำ ๆ ลดหลั่นกัน และทุ่งดอกไม้กว้างสุดตาเผยออกมา กระทั่งรถไฟมาหยุดที่สถานีชนบทที่เงียบเหงา	-1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประโยคที่	ข้อความในประโยค	ค่าอารมณ์
	แห่งหนึ่ง	
23	เจ้าหนูจึงกระโดดลง มันกระโดดลงไปตามขั้นบันไดที่สถานีแล้วมายืนอยู่บนถนนนอกเมืองสายเล็ก ๆ ที่แสนเงียบเหงา	-1
24	ผึ้งส่งเสียงหึ่ง ๆ เบา ๆ ในอากาศ ท้องฟ้าเป็นสีฟ้าเข้มปราศจากเสียงใด ๆ	0
25	ไม่มีภาพผู้คนที่เร่งรีบหรือรถประจำทางเสียงดังให้ได้เห็นเลย	0
26	หนูน้อยดึงบัตรเชิญออกจากกระเป๋าเพื่อจะดูเส้นทางที่ญาติของมันส่งมาให้	0
27	“จากสถานีให้ไปตามถนนสายเล็ก ๆ ลงเนิน แล้วเลี้ยวซ้ายไปทางทุ่งหญ้า”	0
28	มันอ่าน “ตรงไปเรื่อย ๆ จนกว่าจะได้ยินเสียงแม่น้ำ	0
29	แล้วค่อยเลี้ยวขวาที่เดซีดอกที่สาม แล้วเธอจะพบทางเข้าบ้านอยู่ใต้ดอกป๊อปปี้”	0
30	ไม่นานนัก หนูบ้านก็มาเคาะประตูไม้บานเล็ก ๆ ที่อยู่ในร่มเงาของดอกป๊อปปี้ที่กำลังไหวไปมา	0
31	ญาติของมันเปิดประตูออก ร้องจี๊ด ๆ ด้วยความดีใจ “ดีใจจังที่ได้เจอเธอ!”	1
32	หนูนาร้อง “เข้ามาสิ!”	0
33	หนูบ้านก้มหัวลอดเข้าไปข้างใน เช่นเดียวกับบ้านหนูในชนบทหลังอื่น ๆ	0
34	ในบ้านไม่มีอย่างอื่นนอกจากโพรงเล็ก ๆ โพรงหนึ่งในพื้นดิน และผนังว่างเปล่าและพื้นโคลน	0
35	แต่สำหรับหนูบ้านผู้คุ้นเคยกับบ้านหลังใหญ่โต ที่ผนังมีกระดาษปิดตกแต่งและพื้นบ้านที่เงาวับ ที่นี่ก็น่าผิดหวังมากทีเดียว	-1
36	แต่หนูบ้านไม่อยากเสียมารยาท มันจึงแค่พูดออกมาว่า “ทำไมเราไม่ออกไปกินข้างนอกกันล่ะ วันนี้อากาศดีมากเลยนะ!”	1
37	“เป็นความคิดที่ดี” หนูนาพูด	1
38	แล้วมันก็วิ่งพรวดพราดไปที่ห้องเก็บของเพื่อจะเอาอาหารกลางวัน	0
39	ต่อมามันก็กระวีกระวาดออกมาข้างนอกหอบเมล็ดดอกทานตะวันมาเต็มอ้อมแขน	0
40	“ตามสบายเลยนะ” หนูนาพูด	1
41	หนูบ้านแทะเมล็ดพืช “นี่ก็เป็นอาหารธรรมดา ๆ” มันคิด	0
42	“บางทีจานต่อไปอาจจะน่าตื่นเต้นกว่านี้ก็ได้”	1
43	หนูนาสังเกตเห็นว่าญาติของมันกินไปเพียงนิดเดียวจึงพูดขึ้น “นี่จ๊ะ” แล้วก็ยื่นเมล็ดดอกทานตะวันให้หนูบ้านเพิ่มอีก “กินให้เต็มที่เลยนะ”	0
44	“ฉันไม่อยากอิ่มตั้งแต่อาหารเรียกน้ำย่อยน่ะ” หนูบ้านยิ้ม	0
45	“ฉันอยากเก็บท้องไว้เผื่ออาหารหลัก”	0
46	“อาหารหลักงั้นหรือ” หนูนาพูดอย่างแปลกใจ	1
47	“ทั้งหมดก็มีอยู่แค่นี้แหละจ๊ะ”	1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประโยคที่	ข้อความในประโยค	ค่าอารมณ์
48	หนูบ้านประหลาดใจ “แค่นี้เองหรือ” มันร้อง	1
49	“แล้วปกติเธอกินอะไรล่ะ”	0
50	“ก็กินเมล็ดพืชแบบนี้ไง” หนูนาตอบ	0
51	“บางทีก็จะมีสตรอเบอรี่หนึ่งหรือสองลูก แต่ต้องเป็นหน้าของมันนะ...”	0
52	หนูนาหยุดพูดเมื่อเห็นสีหน้าตกใจของญาติ	-1
53	“ไม่มีพายหรือ” หนูบ้านถาม “ไม่มีแซนด์วิช หรือมัฟฟิน หรือชีสบ้างหรือ”	0
54	หนูนาส่ายหัว	0
55	“แล้วเค้กช็อกโกแลต ทาร์ตแยม หรือว่าไอศกรีมล่ะ”	0
56	“ฉันยังไม่เคยชิมด้วยซ้ำ” หนูนาตอบ	0
57	“โถ ญาติที่รักของฉัน” หนูบ้านพูด	1
58	“ชีวิตเธอยังไม่คุ้มค่า ถ้ายังไม่เคยได้ลอง ไอศกรีม”	0
59	หนูบ้านลุกขึ้นยืน แล้วดึงหนูนาให้ลุกตาม	0
60	“เราจะต้องออกจากที่นี่เดี๋ยวนี้” มันพูดอย่างหนักแน่น	1
61	“เธอไปกินข้าวเย็นที่บ้านฉันเถอะ”	0
62	“แต่ว่าฉันยังไม่เคยออกนอกเขตทุ่งนามาก่อนนี่”	0
63	หนูนาร้องเมื่อพวกมันวิ่งไปที่สถานี	0
64	“งั้นก็ได้เวลาที่เธอจะต้องออกจากทุ่งนาแล้วละ” หนูบ้านตอบ	0
65	พวกมันกระโดดขึ้นรถไฟ แล้วปีนป่ายขึ้นไปยังที่นั่งข้างหน้าต่าง	0
66	เมื่อรถไฟค่อย ๆ เคลื่อนตัวผ่านชนบท หนูนามองออกไปนอกหน้าต่างอย่างตื่นเต้น	1
67	จมูกของมันแนบกระจก ภาพทุ่งนาเปลี่ยนเป็นบ้านคนและถนน รถประจำทางและรถยนต์	0
68	ต่อมาไม่นานนักพวกมันก็มาถึงเมืองอันแสนวุ่นวาย	-1
69	เมื่อหนูทั้งสองก้าวออกจากสถานี หนูนาก็อ้าปากกว้าง	0
70	ตัวมันสั่นด้วยความกลัวเมื่อเห็นรถวิ่งอย่างรวดเร็วผ่านหน้ามัน	-1
71	และตัวสั่นมากขึ้นเมื่อรถประจำทางขนาดยักษ์จอดเสียงเอี๊ยดข้างพวกมัน	-1
72	แต่ดูเหมือนหนูบ้านแทบจะไม่ได้สังเกตเห็นด้วยซ้ำ “เร็วเข้า” มันพูด แล้วมุดหายเข้าไปกับการจราจรอันแสนวุ่นวาย	-1
73	กว่าพวกมันจะมาถึงบ้านที่หนูบ้านอาศัย หนูนาก็เกือบจะถูกเหยียบ เฉียดถูกรถทับ และเหน็ดเหนื่อยเป็นที่สุด	-1
74	“ยินดีต้อนรับสู่บ้านอันต่ำต้อยของฉัน” หนูบ้านพูด	1
75	แล้วหนูทั้งสองก็ลอดตัวผ่านประตูหน้าเข้าไป	0
76	หนูนาร้องอุทาน เมื่อเห็นกระเบื้องเป็นมันเงาอยู่ทั่วไปหมด ผนังห้องมีดอกไม้	1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประโยคที่	ข้อความในประโยค	ค่าอารมณ์
	งดงามแปลกตาคลุมอยู่ และบนเพดานมีโคมระย้าทำจากแก้วเจียระไนส่องแสงเหมือนดวงอาทิตย์ห้อยอยู่	
77	“เธอหิวหรือยัง” หนูบ้านถาม	0
78	“หิวตาลายเลยละ” หนูนาตอบ	0
79	“งั้นพวกเรารออะไรกันอยู่ล่ะ” หนูบ้านหัวเราะ	1
80	หนูนาทั้งสองตัววิ่งผ่านพื้นกระเบื้องไปยังห้องอาหาร และปีนขึ้นไปตามขาโต๊ะอาหารขนาดยักษ์	0
81	“อาหารเย็นตั้งโต๊ะแล้ว” หนูบ้านพูดอย่างโอ้อวด	-1
82	ปากของหนูนาอ้ากว้าง	0
83	มันไม่เคยเห็นอาหารมากมายขนาดนี้มาก่อนในชีวิต มีชีสกองเป็นภูเขา และบิสกิตชิ้นเล็ก ๆ เรียงเป็นแถว องุ่นสีม่วงพวงโต และแซนด์วิชที่ชวนให้น้ำลายหกวางเป็นตั้ง ๆ และที่ปลายโต๊ะยังมีเค้กและของหวานน่าอร่อยมากมายเกินกว่าที่มันเคยฝันไว้วางอยู่	1
84	“ตามสบายเลยนะ” หนูบ้านพูด ขณะชีสอยู่เต็มปาก	1
85	หนูนาวิ่งไปรอบโต๊ะ ทำอะไรไม่ถูก	0
86	มันดึงองุ่นตรงนี้แล้วแทะแซนด์วิชตรงนั้น ก่อนจะปีนไปทั่วกองชีส ชิมอาหารตรงนี้นิดตรงนี้หน่อยจนเต็มปาก	0
87	“วี้ด!” หนูบ้านร้องเสียงแหลมจากฝั่งตรงข้ามของโต๊ะ ก่อนจะเลื่อนตัวไปตามช้อนลงมาที่ชามใส่ วิปครีม	0
88	“รอฉันด้วย!” หนูนาตะโกนอย่างร่าเริงแล้วก็วิ่งไปร่วมสนุกด้วย	1
89	หลังจากที่กินเค้กช็อกโกแลตเข้าไป และเลียน้ำตาลบนชิ้นเชอรี่พร้อมกับเคี้ยวทาร์ตสตรอเบอรี่ชิ้นใหญ่อย่างติดหนึบแล้ว หนูนาก็คิดว่าบ้านในเมืองหลังนี้ให้ความรู้สึกเหมือนเป็นบ้านตัวเองเลย	0
90	ขณะที่มันกำลังเช็ดอุ้งเท้าด้วยผ้าเช็ดปากแข็ง ๆ สีขาว หนูบ้านก็ร้องว่า “มากับฉันสิ ฉันมีอะไรให้เธอประหลาดใจ”	1
91	หนูบ้านใช้อุ้งเท้าลากญาติของมันไปยังถ้วยแก้วเย็น ๆ ใบหนึ่งที่มีลูกกลม ๆ ขนาดใหญ่สีขาวและสีชมพูอยู่ในนั้น	0
92	“นี่...” หนูบ้านพูดอย่างตื่นเต้น “ไอศกรีมไงล่ะ”	1
93	หนูทั้งสองแกว่งตัวขึ้นไปบนช้อนแล้วกระโดดลงไปในถ้วยใบนั้น	0
94	“มันเย็นเจี๊ยบเลย!” หนูนาอุทาน	0
95	“แน่อยู่แล้ว” หนูบ้านพูด “ลองชิมดูสิ”	0
96	หนูนาจุ่มอุ้งเท้าลงไปในไอศกรีมแล้วเลีย	0
97	มันส่งความรื่นรมย์อันหอมหวานเย็นจนสั่นสะท้านไปถึงสันหลังของเจ้าหนูน้อย “สวรรค์จริง ๆ เลย!”	1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประโยคที่	ข้อความในประโยค	ค่าอารมณ์
98	หนูนาถอนหายใจ พร้อมกับหลับตาอย่างเป็นสุข	1
99	ทันใดนั้น ประตูห้องอาหารก็เปิดออก ชายคนหนึ่งเดินเข้ามาในห้องเพียงชั่วพริบตา	0
100	หนูบ้านก็กระโดดลงไปซ่อนตัวอยู่ด้านหลังถ้วยไอศกรีม แต่หนูนาตรึงแน่นอยู่กับที่ด้วยความกลัว	-1
101	“หาที่ซ่อนสิ!” หนูบ้านกระซิบ “เร็วเข้า ก่อนที่เขาจะเห็นเจ้า”	0
102	เมื่อสติกลับมา หนูนาก็มุดตัวลงไประหว่างไอศกรีมสองลูก	0
103	มันเย็นกว่าอากาศตอนกลางฤดูหนาวเสียอีก แต่หนูนาก็ไม่กล้าขยับมันหมอบตัวสั่น จนกระทั่งได้ยินเสียงประตูปิดลงอีกครั้ง แล้วเห็นใบหน้าของหนูบ้านกำลังมองลงมา	-1
104	“ตะ ตอน นี้ มะ มัน ปะ ปลอด ภัย หระ หรือ ยะ ยัง” หนูนาถาม ฟันของมันสั่นกระทบกันด้วยความหนาว	-1
105	“ปลอดภัยแล้ว” หนูบ้านตอบ พลางดึงตัวญาติของมันออกมาจากถ้วยไอศกรีม แล้วปัดเกล็ดน้ำแข็งออกจากหนวด	1
106	“มากินพายแอปเปิ้ลกัน” หนูบ้านพูดให้หนูนาคลายความหวาดกลัว “พายจะช่วยให้เธออบอุ่นขึ้น”	-1
107	ขณะที่หนูบ้านและหนูนากำลังมุ่งหน้าไปยังพายแอปเปิ้ลอบที่สุดขอบโต๊ะ ก็มีเสียง “เมี้ยว” ดังขึ้น	0
108	ก่อนที่หนูนาจะทันได้ถามว่าเกิดอะไรขึ้น แมวสีส้มตัวเขื่องท่าทางดุร้ายก็กระโดดขึ้นมาบนโต๊ะตรงหน้าพวกมัน	-1
109	“วิ่งให้เร็วที่สุด” หนูบ้านร้องเสียงแหลม	0
110	คราวนี้ ไม่จำเป็นต้องบอกหนูนาซ้ำสอง ทั้งคู่หนีอย่างอลหม่านโดยมีแมวจี้ตามหลังมาติด ๆ	-1
111	พวกมันวิ่งผ่านเค้กช็อกโกแลต ลอดตัวผ่านช่องระหว่างจานสองใบแล้ววิ่งไปรอบ ๆ ถาดชีส	0
112	แมวขู่อย่างดุร้ายพร้อมกับวิ่งไล่หลังพวกมัน ทำให้แก้วล้มชามปลิวไปทุกทิศทุกทาง	-1
113	หนูทั้งสองวิ่งมาถึงขอบโต๊ะได้ทันเวลาพอดี	0
114	เมื่อหนูนากระโดดลงจากโต๊ะ มันรู้สึกได้ว่ากรงเล็บของแมวพลาดหลังของมันไปอย่างหวุดหวิด	0
115	หนูนากระโดดลงบนพรมผืนนุ่ม หัวใจเต้นถี่รัวด้วยความกลัว ด้านหลังของมัน	-1
116	เจ้าแมวกำลังขู่อย่างโกรธแค้นเมื่อมันเซถลาไปที่ขอบโต๊ะ ตะปบผ้าปูโต๊ะไว้เพื่อกันไม่ให้ตัวเองหล่นลงไป	-1
117	“ทางนี้!” หนูบ้านร้อง ชี้ไปทางรู้เล็ก ๆ ตรงริมผนัง	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประโยคที่	ข้อความในประโยค	ค่าอารมณ์
118	มีเสียงร้องอย่างเจ็บปวดและเสียงโครมครามดังสนั่นเมื่อแมวหล่นลงมาที่พื้น	-1
119	มันดึงผ้าปูโต๊ะและจานกับแก้วทั้งหมดตามลงมาด้วย	0
120	หนูน้อยทั้งสองมุดเข้ารูไปโดยไม่หันกลับมามอง แล้วทรุดตัวลงไปกองที่พื้นด้านหลังกำแพง สีข้างของหนูทั้งสองขยับขึ้นลง	0
121	“นั่นก็แค่ข้อเสียเล็ก ๆ น้อย ๆ ข้อหนึ่งของการใช้ชีวิตแบบคนที่มีระดับน่ะ” หนูบ้านหอบแฮก	-1
122	“ข้อเสียเล็ก ๆ น้อย ๆ งั้นหรือ” หนูนาร้อง	-1
123	“เชิญเธอมีชีวิตแบบมีระดับของเธอต่อไปเถอะ ฉันขอกลับชนบทของฉันดีกว่า”	0
124	“อะไรนะ” หนูบ้านพูด “แล้วก็อยู่โดยไม่มีอาหารจากสวรรค์พวกนี้น่ะหรือ”	0
125	“ใช่” หนูนาตอบอย่างหนักแน่น	1
126	“ฉันยินดีจะอยู่อย่างสงบสุข	1
127	มากกว่ามาเสี่ยงชีวิต เพื่อแลกกับไอศกรีมหนึ่งอุ้งมือแบบนี้”	-1
128	“นิทานเรื่องนี้สอนให้รู้ว่า” ไม่มีที่ไหนจะอุ่นใจ เหมือนบ้านเรา	1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ค.
ผลงานวิจัยที่ได้รับการตีพิมพ์

ภาคผนวกนี้ ว่าด้วยผลงานวิจัยฉบับสมบูรณ์ที่ได้รับการตอบรับ ได้ไปนำเสนอด้วยการพูด (Oral presentation) และได้มีการตีพิมพ์ ซึ่งประกอบไปด้วย

1. **Lertsuksakda, R.**, Netisopakul, P., Pasupa, K. (2014) Thai Sentiment Terms Construction using the Hourglass of Emotions. In: Proceeding of the 6th International Conference on Knowledge and Smart Technology (KST 2014), 30-31 January 2014, Chonburi, Thailand, pp 46-50.
2. **Lertsuksakda, R.**, Pasupa, K., Netisopakul, P., (2015) Sentiment Analysis on Thai Children Stories with Support Vector Machine In: Proceeding of the 20th International Symposium on Artificial Life and Robotics (AROB 2015), 21-23 Jan 2015, Beppu, Japan, pp 138-142

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2014 6th International Conference on Knowledge and Smart Technology (KST)

Thai Sentiment Terms Construction using the Hourglass of Emotions

Rathawut Lertsuksakda
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, Thailand
Email: s3070069@it.kmitl.ac.th

Ponrudee Netisopakul
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, Thailand
Email: ponrudee@it.kmitl.ac.th

Kitsuchart Pasupa
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, Thailand
Email: kitsuchart@it.kmitl.ac.th

Abstract—Most of the current sentiment analysis techniques classifies emotions into two classes which are positive and negative. Some works classify them as positive, negative and objective (neutral). In fact, there are many kinds of emotions in human mind. Recently, psychological viewpoints have influenced most of the works in sentiment analysis. This psychology perspective was adopted to classify human emotions into a wider range, and in a more accurate manner. This paper reviews the adopted computational representation of emotions the so-called Hourglass of Emotion. This paper also proposes a construction of Thai sentiment resource based on such representation for Thai sentiment term tagging. A preliminary sentiment text tagging result shows that the resource as an ontology can be successfully used to tag sentiment text in Thai children stories.

Index Terms—Sentiment analysis, Bidirectional translation, Sentic computing.

I. Motivation

Opinion mining of text content from review product websites is valuable information for both public and product-owner company. In a broader scope, the method is also employed for other domains, such as politics and social responses. One of the most interesting applications is to tag emotions for actors in fiction books, as one proposed by [1]. The work in [2] has identified actors from Thai children story books. However, there remains a work of tagging emotions to each actor speech.

Current state of the art in sentiment analysis either classified sentiments into positive and negative [3]. Some works [4] classified sentiments as positive, negative and objective (or neutral). However, in order to correctly tag emotion from stories or fictions, many emotions must be identified.

Recently, the viewpoint from psychologist has influence the works in sentiment analysis. For example, the wheel of emotions by Robert Plutchik [5] is adopt and adapted into a sentic computing framework, which classified numerous emotions base on four independent but concomitant dimensions [6]. The resource is now available for a public domain [7].

However, for Thai language, there is no such resource exist. Hence, this research proposes to build a Thai sentiment resource from English resource and use it to tag Thai stories.

This paper is organized as follows. Section II reviews the theory of human emotions and other related works. Section III provides details of methodology used. Section IV shows the output implementation results and Section V concludes and discusses future work.

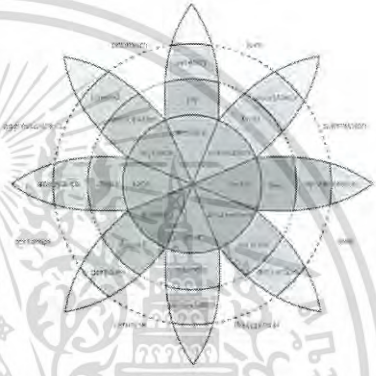


Fig. 1. Plutchik's wheel of emotions. [5]

II. Literature Reviews

A. The Hourglass of Emotions and SenticNet

There were ideas and attempts to develop computational models representing human emotions. Robert Plutchik has invented a wheel of emotions in 1980 [5], as shown in Fig. 1. Plutchik introduced eight basic emotions: i.e. joy, trust, fear, surprise, sadness, disgust, anger, and anticipation. These can be divided into two polarities i.e. joy opposites to sadness, anger opposites to fear, trust opposites to disgust, and surprise opposites to anticipation. There are also three degrees of each emotion, for example, serenity is a lesser degree of joy and ecstasy is a more intense degree of joy.

In addition, Plutchik emotions can combine to form a new emotion. For example, joy and trust are combined to be love. However, joy, trust, and anger are combined to be jealousy.

The hourglass of emotions improved upon Plutchik representation [6]. Instead of classifying each emotion individually,

978-1-4799-1424-1/13/$31.00 ©2013 IEEE

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

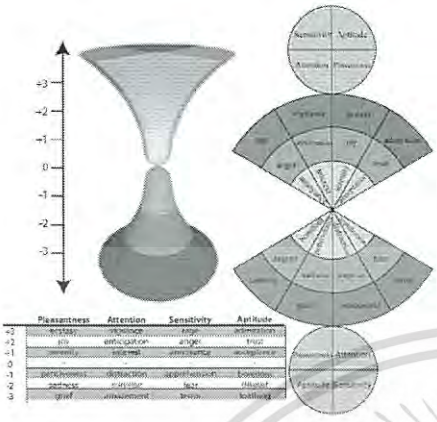


Fig. 2. The hourglass of emotions. [6]

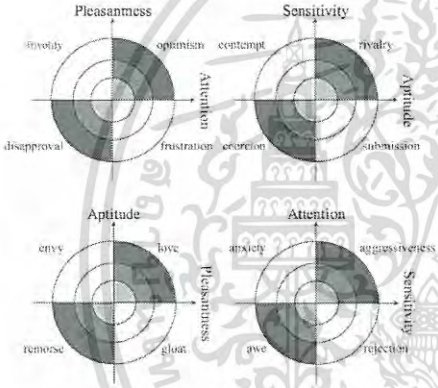


Fig. 3. Illustration of compound emotions. [8]

this model describes each emotion or affective state as four concomitant but independent dimensions. Those are pleasantness, attention, sensitivity and aptitude. Each emotion has ranges of negative values to positive values. Some emotions such as joy, anticipation, anger, and trust have positive values of two but in different dimensions. While the opposite emotion pairs such as joy/sadness, anticipation/surprise, anger/fear and trust/disgust, each has opposite value of two in different direction but the same dimension, as shown in Fig. 2. A vector containing values of the four dimensions for each emotion are called Sentic vector. Fig. 3 demonstrates some second level emotions, such as love has positive degree of attitude and pleasantness; while remorse has negative degrees of both

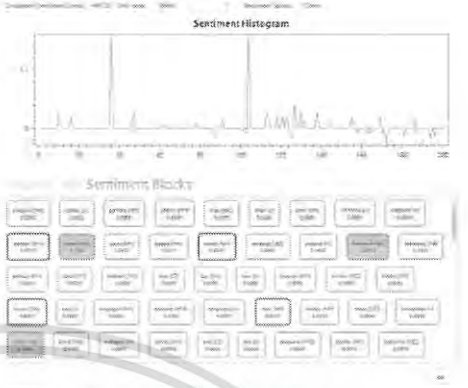


Fig. 4. Example output screenshot from [3]

dimensions.

Although the work by [4] attempted to create a knowledge base for human emotions, more advanced work in sentiment analysis can be found in [8,9]. The term sentic computing is founded in [8] and is described as a multi-disciplinary approach to sentiment analysis that exploits both computer and social sciences to better recognize, interpret, and process opinions and sentiments over the Web. The multi-disciplinary includes Artificial Intelligence (AI) and semantic web techniques, mathematics, linguistics, psychology, sociology and ethics. This derives into a set of sentiment resources and tools for processing English sentiment analysis.

SenticNet2 [9] is a product of sentic computing available thru [7]. This semantic and affective resource for sentiment analysis is employed by other researchers such as [10].

Other works such as pSenti [3], which did not employ SenticNet, only classified emotions into two polarities (positive and negative). Their work used lexical analysis to build a basic set of sentiment terms as a knowledge base. The intense of sentiments in an input text are raised or lowered by various adjective terms. Moreover, a supervised machine learning algorithm, LibSVM, is used to adjust the weight of lexical terms in order to provide more accurate sentiment values of input text. The example output screenshot from [3] is shown in Fig. 4.

B. Bidirectional Translation

Bidirectional translation is a two-step process, beginning by translating a source term into target terms using a source-target dictionary. Then translate each target term back to the source language using target-source dictionary. Only target terms that correspond to the original source term are kept. The motivation behind this approach is to find the more exact match meaning of the source term and eliminate vague or spurious terms. The method was used in [11].

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

III. Methodology

A. Thai Sentiment Terms Construction

This research employs a SenticNet2 resource [7], which contains English sentiment terms associated with their sentic values. The sentic values are modeled using The Hourglass of Emotions as described above.

The sentic data is in an RDF file format. In addition, APIs are provided through these URL.

- http://sentic.net/api/en/concept
- http://sentic.net/api/en/concept/CONCEPT_NAME
- http://sentic.net/api/en/concept/CONCEPT_NAME /semantics
- http://sentic.net/api/en/concept/CONCEPT_NAME /sentics
- http://sentic.net/api/en/concept/CONCEPT_NAME /polarity

To obtain Thai terms which match closely to English terms. We employ a bidirectional translation method using four English-Thai and Thai-English dictionary. Those are:

- NECTEC LEXiTRON English-Thai Dictionary
- NECTEC LEXiTRON Thai-English Dictionary
- VOLUBILIS English-Thai Dictionary
- VOLUBILIS Thai-English Dictionary

A bidirectional translation program is developed using Python and pystardict library. It works with an electronic dictionary in a StarDict Dictionary format [12]. The structure of a StarDict Dictionary consists of three following files.

- DICTIONARY_NAME.ifo
- DICTIONARY_NAME.idx
- DICTIONARY_NAME.dict

The sentic values and polarity value can be obtained from the original English term. However, one Thai term might correspond to more than one English term. In this case, the value of the target term is an average of those original terms. The example is provided in the next section.

B. Ontology Design

Our designed ontology as shown in Fig. 5 has two main classes.

1) Concept class: contains four types of vocabularies which are nouns, verbs, adjectives, and adverbs.

2) ValuePartition class: has two subclasses which are the SenticValue subclasses, further divided into the four subclasses, and the PolarityValue subclass, in which the value can be derived from the other subclasses.

C. Sentiment Text Tagging System (STTS)

Sentiment Text Tagging System (STTS) is developed by Python [13]. It aims to tag at both word and sentence levels using the four sentic dimensions described above. The architecture of the STTS is shown in Fig. 6. There are four main components.

An user interface component is for a user to invoke the system and for the system to display the sentiment text tagging results. A Natural Language Processing (NLP) component

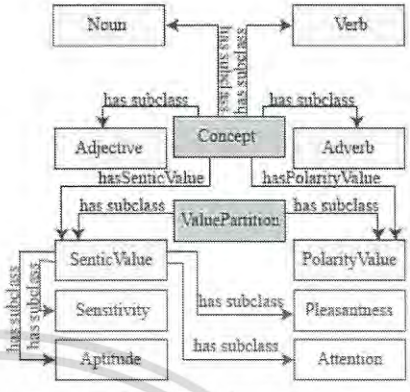


Fig. 5. The designed ontology.

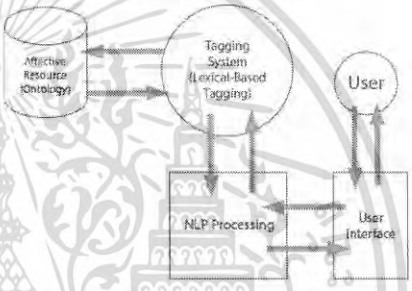


Fig. 6. Architecture of Sentiment Text Tagging System.

is used to preprocess the input text, such as Thai word segmentation. This is necessary for Thai language because the STTS work at both word level and sentence level. A tagging process component is a main working component of the STTS. However, for this preliminary version of the STTS, it only uses to test the feasibility of a new constructed affective resource, which is in a form of ontology. Hence, the process works by matching input text to the vocabulary in the resource. Other functions, which enhance the capability of the system, will be implemented in the future work. An affective resource is stored as an ontology, in which its instances derived from the bidirectional translation method using the SenticNet2 [9] as described above.

IV. Implementation Results

A. English to Thai bidirectional translation results

1) English to Thai bidirectional translation using LEXiTRON dictionary: A LEXiTRON dictionary is a public domain dictionary available at [14]. The SenticNet2 resource [7] contains 14,244 terms, only 8,793 of them has Thai meaning

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

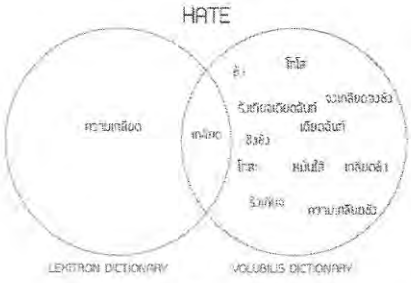


Fig. 7. The result of translating the term hate.

TABLE I
SENTIC VALUES OF 'โกรธ'

Concept	Pleasantness	Attention	Sensitivity	Aptitude
Wrath	-0.737	0	0.899	0
Hate	-0.714	0	0.83	-0.962
Hatred	-0.125	0.121	0.446	-0.652
Fury	0	0	0.786	-0.965
Rage	0	0	0.859	0
Animosity	-0.989	0	0.871	-0.976
Anger	0	0	0.95	0
Average	-0.366	0.017	0.806	-0.508

in the English-Thai LEXiTRON dictionary. Using bidirectional translation, there remains only 1,315 English terms with Thai translation. These correspond to 1,440 Thai terms.

2) English to Thai bidirectional translation using VOLUBILIS: Another public domain English to Thai dictionary is VOLUBILIS [15]. From 14,244 English terms in the SenticNet resource, there are 8,728 English terms with Thai translation. When applying the bidirectional translation, it results in only 4,688 English terms with Thai translation, corresponding to 16,478 Thai terms.

3) The union of both bidirectional translation results: The union of 1,440 Thai terms from LEXiTRON and 16,478 Thai terms from VOLUBILIS has resulting in 16,584 Thai terms after removing redundancy terms. Hence, this set can be used for the sentiment text tagging system.

For example, Fig. 7 shows the result of translating the term hate, which only has two translation in LEXiTRON (ความเกลียด, เกลียด), but many translations in VOLUBILIS (จงเกลียดจงชัง, ชิงชัง, รังเกียจ, โกรธ, เกลียด, หมั่นไส้, ความเกลียดชัง, เดียดฉันท์, รังเกียจเดียดฉันท์, โกโธ, ชัง และเกลียดชัง).

After merging the results from bidirectional translation, we found that one Thai term may match more than one English term. In this case, the four sentic values of the Thai term can be calculated using average values of those English terms. For example, the term 'โกรธ' in Thai can be translated as wrath, hate, hatred, fury, rage, Animosity, and Anger in English. Hence, the sentic values of 'โกรธ' are the average values of those terms, as shown in Table I.

Hence, the term โกรธ has Pleasantness, Attention, Sensitivity, Aptitude values as -0.366, 0.017, +0.446, -0.652, respectively

B. Ontology Development Result

The ontology is developed using Protégé tool [16]. The class structure, object properties and datatype properties in Fig. 8, Fig. 9 and Fig. 10 respectively.

After the ontology construction is done, the system instantiates Thai sentiment terms and their sentic values into the ontology thru Protégé API.

C. Experimenting STTS

Example input and output is shown in Table II and Table III.

Fig. 8. Ontology classes.

Fig. 9. Object properties.

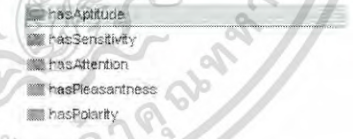


Fig. 10. Datatype properties.

To interpret the sentimental result of the Thai sentence in Table III, the Attention dimension has high positive value and the Sensitivity dimension has high negative value. This shows alertness or suspicious for incoming danger, which is accurate comparing to human reading.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

TABLE II
EXCERPT OF THAI TEXT INPUT FROM A CHILDREN STORY

หนูน้อยหมวกแดงตกใจกลัวจนเกือบจะทำตะกร้าหลุดมือ

TABLE III
OUTPUT OF SENTIMENT TEXT TAGGING SYSTEM

Concept	Pleasantness	Attention	Sensitivity	Aptitude
หนู	0	0	0	0
น้อย	0	0.079	0.097	0.168
หมวก	0	0	0	0
แดง	-0.076	0.101	0	0.105
ตกใจ	-0.467	0.346	-0.216	-0.183
กลัว	0	0	-0.887	-0.398
จน	0	0.118	0	0.115
เกือบ	0.07	0.214	-0.086	0.264
จะ	-0.059	0.153	-0.046	0.158
ทำ	-0.077	0.099	-0.113	0.121
ตะกร้า	0	0	0	0
หลุดมือ	0	0	0	0
Total	-0.609	1.11	-1.251	0.35

However, note that for a word level, many neutral terms, such as เกือบ จะ ทำ should not contribute to the overall emotion. In addition, the term such as น้อย and จน in this context, should not contribute to the overall emotion of this sentence. Therefore, the future work should resolve these issues.

V. Conclusion and Future Work

This paper proposes a methodology to construct a Thai sentiment resource based on sentic computing. The implementation is achieved using pystardict library in Python. From initial 14,244 English terms, the remained bidirectional translation has only 4,708 English terms corresponding to 16,584 Thai terms. The resulting resource is used to tag Thai children story. The preliminary outcome is satisfied.

Nevertheless, the initial investigation found that, at a word level sentiment tagging, some words should not contribute sentiment values to the overall sentence. This issue deserves more investigation in the future.

References

[1] P. Netisopakul, K. Wohlgenannt, S. Wangsiripitak, and K. Pasupa, "Emotional speech synthesis for visibility impaired: From book-to-speech (essvibs)," *Research Project Funding Application submitted to National Research Council of Thailand*, 2012.

[2] P. Netisopakul and P. Wikaha, "Speaker identification from multiple short stories without an avatar list," *The Tenth Symposium on Natural Language Processing (SNLP-2013)*, 2013.

[3] A. Mudinas, D. Zhang, and M. Levene, "Combining lexicon and learning based approaches for concept-level sentiment analysis," *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*, no. 5, pp. 1–8, 2012.

[4] S. Baccianella, A. Esuli, and F. Sebastiani, "Proceedings of the seventh conference on international language resources and evaluation (lrec10)," *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*, pp. 2200–2204, 2010.

[5] R. Plutchik, *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion 1*. New York: Academic Press, 1980.

[6] E. Cambria, A. Livingstone, and A. Hussain, "The hourglass of emotions," *Cognitive Behavioral Systems*, vol. 7403, p. 144157, 2012.

[7] (2013, Sep.) Senticnet 2 @ONLINE. [Online]. Available: http://sentic.net/

[8] E. Cambria and A. Hussain, *Sentic computing: Techniques, tools, and applications*. Berlin Heidelberg: Springer, 2012.

[9] E. Cambria, C. Havasi, and A. Hussain, "Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis," *FLAIRS*, 2012.

[10] K. Dinakar, B. Jones, C. Havasi, H. Lieberman, and R. Picard, "Common sense reasoning for detection, prevention, and mitigation of cyberbullying," *ACM Trans*, 2012.

[11] J. Wan and D. Oard, "Combining bidirectional translation and synonymy for cross-language information retrieval," *Proceedings of the 29th Annual SIGIR Conference*, 2006.

[12] StarDict Team, Inc, "Stardict [online]," 2013. Available: http://www.stardict.org/.

[13] Python Software Foundation, "Python 2.7.3 [online]," 2013, http://www.python.org/.

[14] National Electronics and Computer Technology Center, "Lexitron dictionary [online]," 2013, http://lexitron.nectec.or.th/.

[15] Belisan, "Volubilis dictionary [online]," 2013, http://belisan-volubilis.blogspot.be/.

[16] Stanford Center for Biomedical Informatics Research, "Protege open source ontology editor and knowledge-base framework [online]," 2013, http://protege.stanford.edu/.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

PROCEEDINGS OF THE TWENTIETH INTERNATIONAL SYMPOSIUM ON ARTIFICIAL LIFE AND ROBOTICS

(AROB 20th 2015)

AROB 20th ANNIVERSARY

January 21– 23, 2015

B-Con Plaza, Beppu, Oita, JAPAN

Publication Date: January 21, 2015
ISBN 978-4-9907582-1-9
ISSN 2185-3797

International Society of Artificial Life and Robotics

助成　日本万国博覧会記念基金
Supported by the Japan World Exposition 1970 Commemorative Fund.
この助成金は、日本万国博覧会の収益を基にしています。
公益財団法人　関西・大阪21世紀協会

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The Twentieth International Symposium on Artificial Life and Robotics 2015 (AROB 20th 2015),
B-Con Plaza, Beppu, Japan, January 21-23, 2015

Sentiment analysis of Thai children stories on support vector machine

Rathawut Lertsuksakda[1], Kitsuchart Pasupa[2], Ponrudee Netisopakul[3]

Knowledge Management and Knowledge Engineering Laboratory,
Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Thailand

[1]rathawut.l@gmail.com, [2]kitsuchart@it.kmitl.ac.th, [3]ponrudee@it.kmitl.ac.th

Abstract: Recently, Sentiment Text Tagging System (STTS) with Thai sentiment resource has been developed and used to tag emotions directly to words and sentences in Thai children stories. The Thai sentiment resource, developed from SenticNet2 resource, groups emotions into four independent but concomitant dimensions: pleasantness, attention, sensitivity and aptitude. The measure of each dimension is called a sentic value of that dimension. Thai sentiment resource stores each words sentic value and polarity value, a value calculated from the sentic value, in the form of floating point number. The resource was constructed from bi-directional translation of 14,244 English terms in SenticNet2 into 16,584 Thai terms. The main purpose of this study was to implement a sentiment analysis of Thai children stories system with support vector machine using a set of proposed discriminating features for classifying emotions. It was found that the system can achieve 72.14% of accuracy.

Keywords: SenticNet2; sentiment analysis; emotion analysis; support vector machine.

1 INTRODUCTION

Natural Language Processing (NLP) is a field of study which involves with the interactions between human natural languages and computer. Sentiment analysis is one of the research fields in NLP. It aims to determine the emotions of the sentences or writers. There are the benefits of sentiment analysis, such as the use of sentiment analysis in social monitoring for branding and markerting and the use of sentiment analysis in humanoid robot. Many frameworks of sentiment analysis in English have been implemented, for example, pSenti [1] and sentiment analysis in Twitter task [2].

This paper focuses on Thai sentiment analysis. The sentiment analysis in Thai language is a challenging task because Thai language is not fully equipped with NLP tools like English language. It is known that there is neither explicit sentence marker nor word boundary in Thai language while there is a full stop marking as a sentence marker and white space between two consecutive words as a word boundary in English. One of the recent works is the S-Sense [3]. It is a Thai sentiment analysis system which is introduced by National Electronics and Computer Technology Center (NECTEC) focusing on intention and sentiment analysis in Thai. This framework uses the Naive Bayes to analyze four different intentions i.e. announcement, request, question and sentiment but only two sentiments are considered i.e. positive and negative. Recently, Sentiment Text Tagging System (STTS) with Thai sentiment resource has been proposed and used to tag emotions directly to words and sentences in Thai children stories [4]. Thai sentiment resource based on the hourglass of emotions was constructed from the bidirectional translation method on the SenticNet2 resource [5].

As mentioned above that the S-Sense only consider two sentiments, the sentiment structures are more complex in real world application as there are many type of sentiments. Therefore, this work will focus on a more complex problem with three types of sentiments (i.e. positive, negative and neutral) than the well-known implementations of binary classification (i.e. positive, and negative). We apply a well-known machine learning technique the so-called support vector machine algorithm (SVM) in a conjunction with one-vs-all multiclass scheme to tackle the problem. A set of discriminating features is proposed and analysed on Thai children stories.

This paper is organized as follows: Section 2 outlines the methodologies and tools used in this work. Section 3 gives the dataset details and feature extraction while Section 4 explains experimental framework and followed by experiment results in Section 5.

2 METHODOLOGY

2.1 The Hourglass of Emotions

An affective categorisation model called the Hourglass of Emotions (as shown in Fig. 1) was inspired by a Robert Plutchiks studies on human emotions [6]. Robert Plutchik has invented a wheel of emotions that has 8 primary bipolar emotions (i.e. joy–sadness, anger–fear, trust–disgust and surprise–anticipation) [7].

The Hourglass of Emotions reinterprets a wheel of emotions and classified human emotions into four independent but concomitant dimensions (i.e. pleasantness dimension, attention dimension, sensitivity dimension and aptitude dimension). Each dimension has two directions which are positive and negative direction. The four dimensions are represented as sentic values and provides the polarity value which can be calculated by

$$P = \sum_{i=1}^{N} \frac{Pleasantness(e_i) - Attention(e_i) - |Sensitivity(e_i)| + Aptitude(e_i)}{3N}$$

where N is the number of words.

2.2 SenticNet2

SenticNet2 is a collection of commonly used polarity concepts. It is a semantic and affective resource for opinion

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The Twentieth International Symposium on Artificial Life and Robotics 2015 (AROB 20th 2015),
B-Con Plaza, Beppu, Japan, January 21-23, 2015

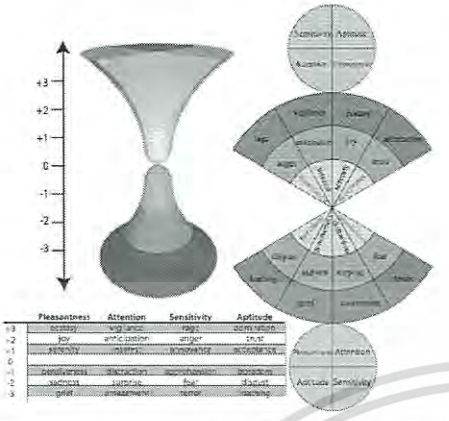


Fig. 1. The Hourglass of Emotions. [6]

mining and sentiment analysis. The collection provides sentic values and polarity value that were performed through the combination of affective space and the hourglass of emotions. It contains 14,244 English polarity concepts and was kept in an RDF file format. The APIs of the resource are available on http://sentic.net/api/en/concept.

2.3 Thai sentiment resource from Sentiment Text Tagging System

The Thai sentiment resource from Sentiment Text Tagging System (STTS) [4] employs a SenticNet2 resource. The system contains 16,584 Thai polarity concepts resulting from the bidirectional translation method onto 14,244 English polarity concepts. The sentic values are modelled using the hourglass of emotions as described above.

2.4 KUCut

It is the word segmentor written in Python programming language for Thai language. The performance of this tool was proved to be more accurate than the previous tool (i.e. CTTEX, SWATH). KUCut is available available on [8].

2.5 Jitar

Jitar is a Thai part-of-speech (POS) tagger of NAiST corpus composing of 17 part-of-speech groups and 48 part-of-speech tags as shown in table 1. Jitar is written in Java programming language.

2.6 Support Vector Machine

Support Vector Machine (SVM) is a powerful supervised learning model which uses support vectors to represent the examples as points in space. It was invented by first by Vapnik in 1995 [9].

Training data with discriminating features and the target class will be analysed to find a hyperplane that separates these training data points into two classes. This hyperplane will be used to separate the test data with discriminating

Table 1. Jitar POS groups, tags and the frequency of each POS that were marked to be the most affective word in the 40 Thai children stories dataset.

POS groups	POS tags	*f*
1. NOUN	1) npn (Proper noun)	20
	2) nnum (Cardinal number)	-
	3) norm (Ordinal Number Marker)	-
	4) nlab (label noun)	-
	5) ncn (Common Noun)	73
	6) nct (Collective Noun)	-
	7) ntit (title Noun)	-
2. PRONOUN	8) pper (personal Pronoun)	-
	9) pdem (Demonstrative Pronoun)	-
	10) pind (indefinite Pronoun)	-
	11) ppos (Possessive Pronoun)	-
	12) prfx (Reflexive Pronoun)	-
	13) prec (Reciprocal Pronoun)	-
	14) prel (Relative pronoun)	-
	15) pint (Interrogative Pronoun)	-
3. VERB	16) vi (intransitive Verb)	518
	17) vt (Transitive Verb)	268
	18) vcau (Causative Verb)	-
	19) vcs (Complementary State Verb)	-
	20) vex (Existential Verb)	-
	21) prev (Pre-Verb)	1
	22) vpost (Post-verb)	-
	23) hoam (honorific marker)	-
4. DETERMINER	24) det (determiner)	-
	25) indet (Indefinite determiner)	-
5. ADJECTIVE	26) adj (Adjective)	60
6. ADVERB	27) adv (adverb)	76
	28) advm1 (Adverb Marker1)	-
	29) advm2 (Adverb Marker2)	-
	30) advm3 (Adverb Marker3)	-
	31) advm4 (Adverb Marker4)	-
	32) advm5 (Adverb Marker5)	-
7. CLASSIFIER	33) cl (Classifier)	-
8. CONJUNCTION	34) conj (Conjunction)	1
	35) conjd (Double Conjunction)	-
	36) conjncl (Noun Clause Conjunction)	1
9. PREPOSITION	37) prep (Preposition)	-
	38) prepc (co-Preposition)	-
10. INTERJECTION	39) int (interjection)	1
11. PREFIX	40) pref1 (Prefix1)	-
	41) pref2 (Prefix2)	-
	42) pref3 (Prefix3)	-
12. PARTICLE	43) aff (Affirmative)	2
	44) part (particle)	1
13. NEGATIVE	45) neg (Negative)	-
14. PUNCTUATION	46) punc (punctuation)	-
15. IDIOM	47) idm (IDIOM)	1
16. PASSIVE VOICE MARKER	48) psm (Passive Voice Marker)	1
17. SYMBOL	49) sym (Symbol)	-
ALL		1024

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The Twentieth International Symposium on Artificial Life and Robotics 2015 (AROB 20th 2015),
B-Con Plaza, Beppu, Japan, January 21-23, 2015

features in order to predict the class of each test sample.

3 THE 40 THAI CHILDREN STORIES DATASET

The dataset was collected from 40 Thai children stories [10]. We manually split the text into sentences which leads to 1,964 sentences. Each sentence was manually tagged by human into three classes which are positive, negative and neutral according to the reader's expressed emotion. This can be divided into 573 positive class (29.18%), 940 neutral class (47.86%), and 451 negative class (22.96%).

KUCut was used to carry out the word segmentation of sentences in dataset while part-of-speech was tagged by Jitar. STTS was used to tag four main sentic values directly to words. Each sentence is represented by a set of discriminating features which are polarity and sentic values of each concomitant sentiment: pleasantness, attention, sensitivity and aptitude. These features are extracted based on the number of selected words with the highest amplitude of polarity value in a sentence. The word is selected if their POS is either vi, vt, adv, ncn or adj. While the users were tagging each sentence, they were also asked to mark the most affective word which stimulated their emotion. Table 1 shows the frequency of each POS that were marked to be the most affective word. The selected POSs are the top five of ranked frequency. This is the reason behind the selection of this set of POS.

In this paper, we considered the sentic values and polarity value of the top one, two, three, four, five or six words with the highest amplitude of polarity value. It should be noted that these words belong to the set of selected POS. A simple example of feature extraction procedure is shown in Fig. 2. The input is "การเคลื่อนไหวของต้นอ้อสวยมากเลยนะ" that means "The movement of the reed is very beautiful". Therefore, the feature vector of this input sentence is [0.29, 0.35, -0.07, 0.38, 0.32, 0.21, 0.18, -0.09, 0.34, 0.21].

4 EXPERIMENTAL FRAMEWORK

We evaluate the proposed discriminating feature set using SVM. Three type of kernels used in this experiment are linear, radial basis function (RBF), and polynomial kernel. In the experiment, ten-fold cross validation is used to identify the optimal tuning parameters e.g. regularisation (C) and kernel parameter for RBF and polynomial function). A set of optimal parameters is selected based on the accuracy. In linear case, there is only "C" parameter to be tuned while RBF and polynomial has an additional optimal kernel parameter to be obtained. The range of the C parameter is $10^{-6} - 10^{4}$. The range of kernel parameter for RBF and polynomial function is $10^{-6} - 10^{7}$ and $2 - 10$, respectively. The optimal tuning parameters are used to train the classifiers for each type of kernel. It should be noted that we used one-vs-all multiclass classification scheme to tackle the problem.

The 1,964 sentences were randomly divided in half to training and test set 1,000 times in order to ameliorate sample bias. Each variable is normalized to zero mean and unit standard. Prior distribution of each class is maintained for both training and test set.

5 EXPERIMENT RESULTS

The results reported here are averaged across these 1,000 sets. Table 2 reports the accuracy of SVM for each kernel together with its optimal parameters. It is found that SVMs in a conjunction with all type of kernel function are better than the baseline (47.86%). The baseline is equal to the prior distribution of the major class (here, neutral class) as the classifier predicts a single label for all samples. RBF with two selected words in a sentence gives the best result at 72.14% of accuracy followed by polynomial and linear function with three selected words in a sentence.

Table 2. Accuracy of SVM with each type of kernel (average across 1,000 random splits).

Number of Selected Words	Kernel	C	Kernel Parameter	Accuracy (%)
1	Linear	0.01	-	59.36 ± 3.39
2	Linear	10.00	-	64.16 ± 1.17
3	Linear	100.00	-	**64.99 ± 1.21**
4	Linear	0.10	-	64.80 ± 1.18
5	Linear	0.10	-	64.61 ± 1.16
6	Linear	10.00	-	64.56 ± 1.15
1	RBF	0.10	1.00	68.99 ± 1.14
2	RBF	1.00	0.10	**72.14 ± 1.09**
3	RBF	1.00	0.10	71.17 ± 1.11
4	RBF	1.00	0.10	70.14 ± 1.14
5	RBF	1.00	0.10	69.16 ± 1.21
6	RBF	10.00	0.01	68.27 ± 1.14
1	Polynomial	10.00	7.00	63.63 ± 1.21
2	Polynomial	1,000.00	5.00	64.65 ± 1.27
3	Polynomial	100.00	2.00	**65.01 ± 1.24**
4	Polynomial	1.00	2.00	63.26 ± 1.25
5	Polynomial	10.00	2.00	62.38 ± 1.26
6	Polynomial	10.00	2.00	61.93 ± 1.26

The results in Table 2 are also tested by independent two-sample t-test in order to determine whether there were significant difference between the mean values obtained from each type of kernel. The results of t-test are reported in Table 3. It can be seen than the results are significant at $p < 0.001$, except the case of linear and polynomial kernel with three feature words are less conclusive ($p = 0.7599$).

Fig. 3-5 shows the receiver operating characteristic (ROC) curve of SVM in a conjunction with linear (three feature words), RBF (two feature words), polynomial (three feature words) kernel function, respectively. These graphs are averaged across 1,000 random splits. All the classifiers are better than random baseline because all the points are above the diagonal, which represent good classification results. Again, SVM-RBF can achieve the highest average area under the ROC curve at 0.84.

We also show the confusion matrix of SVM in a conjunction with linear, RBF, polynomial kernel function, in Fig. 6-8, respectively. These value are average numbers across 1,000 random splits. All correct predicted are located in the diagonal of the matrix, and the errors are represented by values outside the diagonal. The results show that SVM-RBF can achieve higher correct predicted rate on the minority classes (positive and negative) than SVM-Linear and SVM-Polynomial.

©ISAROB

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The Twentieth International Symposium on Artificial Life and Robotics 2015 (AROB 20th 2015),
B-Con Plaza, Beppu, Japan, January 21-23, 2015

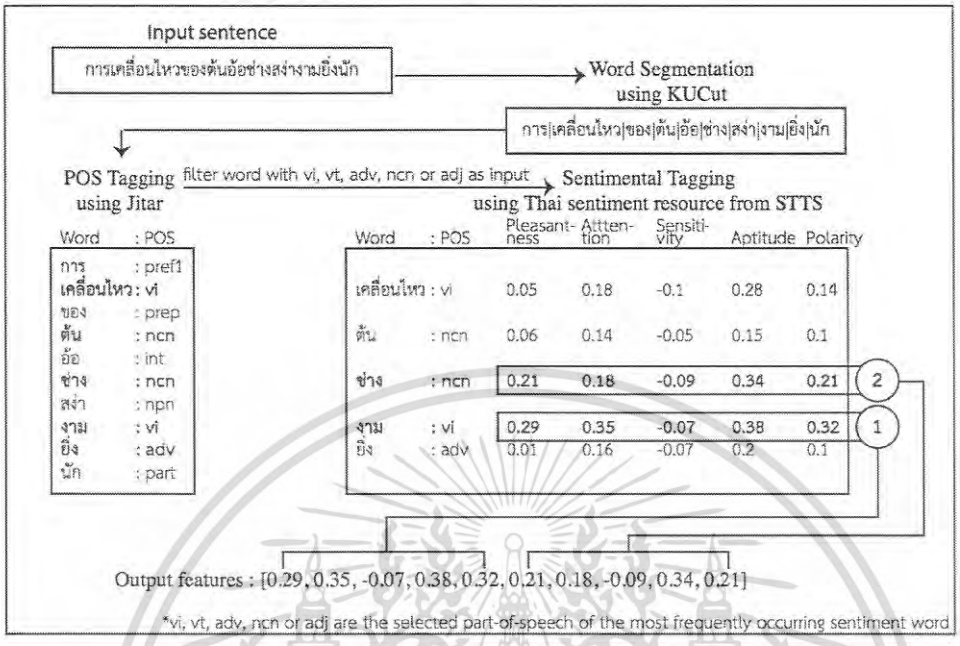


Fig. 2. A simple example of features extraction when two words (highest polarity values) are selected.

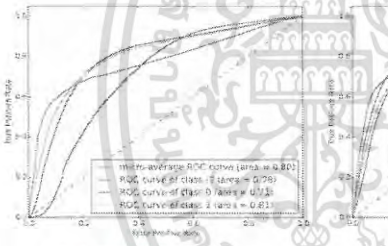


Fig. 3. ROC curve of SVM-Linear

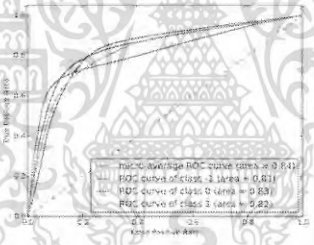


Fig. 4. ROC curve of SVM-RBF

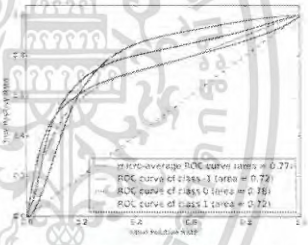


Fig. 5. ROC curve of SVM-Polynomial

6 CONCLUSION

As above-mentioned, the main purpose of this study was to implement a sentiment analysis of Thai children stories on a support vector machine (SVM) using a set of proposed discriminating features for classifying emotions. Several analysis runs with different SVM's kernels at various parameter settings were performed, and the results were compared. It was found that the best accuracy was obtained with the following SVM in a conjunction with RBF kernel and using the sentic and polarity values of the top two words with the highest polarity amplitude.

References

[1] A. Mudinas, D. Zhang, and M. Levene, "Combining lexicon and learning based approaches for concept-level sentiment analysis," in *Proceedings of the 1st International Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM 2012), 12-16 August 2012, Beijing, China*, 2012, pp. 1–8.

[2] S. Rosenthal, A. Ritter, P. Nakov, and V. Stoyanov, "Semeval-2014 task 9: Sentiment analysis in twitter," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), 23-24 August 2014, Dublin, Ireland*, 2014, pp. 73–80.

[3] C. Haruechaiyasak, A. Kongthon, P. Palingoon, and K. Trakultaweekoon, "S-sense: A sentiment analysis

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

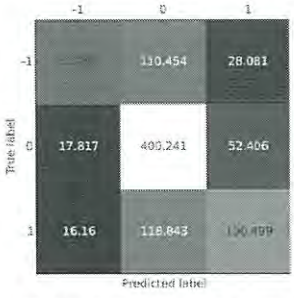


Fig. 6. Confusion matrix of SVM-Linear

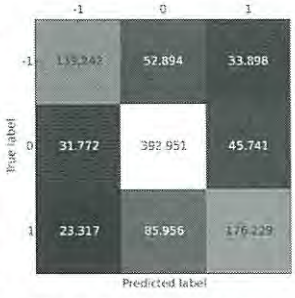


Fig. 7. Confusion matrix of SVM-RBF

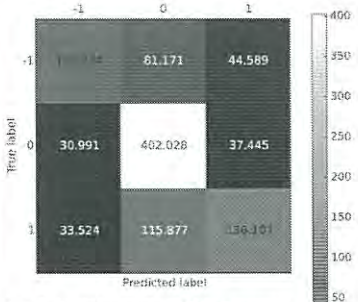


Fig. 8. Confusion matrix of SVM-Polynomial

Table 3. The results of independent two-sample t-test.

Number of Feature Words	Kernel-1	Kernel-2	t-statistic	p-value
1	Linear	RBF	-84.98	< 0.001
1	Linear	Polynomial	-37.51	< 0.001
1	RBF	Polynomial	101.65	< 0.001
2	Linear	RBF	-157.83	< 0.001
2	Linear	Polynomial	-8.92	< 0.001
2	RBF	Polynomial	141.69	< 0.001
3	Linear	RBF	-119.18	< 0.001
3	Linear	Polynomial	-0.31	0.7599
3	RBF	Polynomial	117.34	< 0.001
4	Linear	RBF	-102.83	< 0.001
4	Linear	Polynomial	28.28	< 0.001
4	RBF	Polynomial	128.65	< 0.001
5	Linear	RBF	-85.72	< 0.001
5	Linear	Polynomial	41.14	< 0.001
5	RBF	Polynomial	122.70	< 0.001
6	Linear	RBF	-72.18	< 0.001
6	Linear	Polynomial	48.84	< 0.001
6	RBF	Polynomial	117.91	< 0.001

framework for social media sensing," in *Proceedings of the IJCNLP 2013 Workshop on Natural Language Processing for Social Media (SocialNLP), 14 October 2013, Nagoya, Japan*, October 2013, pp. 6–13.

[4] R. Lertsuksakda, P. Netisopakul, and K. Pasupa, "Thai sentiment terms construction using the hourglass of emotions," in *Proceeding of the 6th International Conference on Knowledge and Smart Technology (KST 2014), 30-31 January 2014, Chonburi, Thailand*, 2014, pp. 46–50.

[5] E. Cambria, C. Havasi, and A. Hussain, "Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis." in *Proceeding of the 25 International Florida Artificial Intelligence Research Society (FLAIRS 2012), 23-25 May 2012, Marco Island, Florida, USA*, 2012, pp. 202–207.

[6] E. Cambria, A. Livingstone, and A. Hussain, "The hourglass of emotions," in *Cognitive Behavioural Systems, Lecture Note in Computer Science*, 2012, vol. 7403, pp. 144–157.

[7] R. Plutchik and H. Kellerman, *Emotion: Theory, Research and Experience. Vol. 1, Theories of Emotion*. Academic Press, 1980.

[8] S. Sudprasert, "KUCut thai word segmentor [online]," 2004, http://belisan-volubilis.blogspot.be/.

[9] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[10] P. Netisopakul and P. Wikaha, "Speaker identification from multiple short stories without an avatar list," in *Proceeding of the 10th Symposium on Natural Language Processing (SNLP-2013), 28-30 October 2014, Phuket, Thailand*, 2013, pp. 1–7.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ผู้เขียน	นายรัฐวุฒิ เลิศสุขศักดา
วันเดือนปีเกิด	18 มีนาคม 2535
สถานที่เกิด	กรุงเทพมหานคร
ปริญญา	2557 วิทยาศาสตรบัณฑิต สาขาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง 2558 วิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ แขนงวิชาวิทยาการสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
งานวิจัยที่ตีพิมพ์	**Lertsuksakda, R.**, Netisopakul, P., Pasupa, K. (2014) Thai Sentiment Terms Construction using the Hourglass of Emotions. In: Proceeding of the 6th International Conference on Knowledge and Smart Technology (KST 2014), 30-31 January 2014, Chonburi, Thailand, pp 46-50. **Lertsuksakda, R.,** Pasupa, K., Netisopakul, P., (2015) Sentiment Analysis on Thai Children Stories with Support Vector Machine In: Proceeding of the 20th International Symposium on Artificial Life and Robotics (AROB 2015), 21-23 Jan 2015, Beppu, Japan, pp 138-142

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้