



โครงการ

การเรียนการสอนเพื่อเสริมประสบการณ์

ชื่อโครงการ การวิเคราะห์ความรู้สึกจากอิโมจิเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์
Twitter User's Sentiment Analysis from Emoji on Tourism

ชื่อนิสิต นางสาวนกกานต์ ต่านสว่างกุล 583 36030 23

ภาควิชา คณิตศาสตร์และวิทยาการคอมพิวเตอร์
สาขาวิชา วิทยาการคอมพิวเตอร์

ปีการศึกษา 2561

บทคัดย่อและไฟล์ฉบับสมบูรณ์ของโครงงานทางวิชาการที่ให้บริการในคลังข้อมูลความรู้ (CUIR)
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
เป็นแฟ้มข้อมูลของนิสิตเจ้าของโครงการทางวิชาการที่ส่งผ่านทางคณะที่สังกัด

The abstract and full text of senior projects in Chulalongkorn University Intellectual Repository(CUIR)
are the senior project authors' files submitted through the faculty.

การวิเคราะห์ความรู้สึกจากอีโมจิเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์

นางสาวกนกกานต์ ด้านสว่างกุล

โครงการนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2561
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

TWITTER USER'S SENTIMENT ANALYSIS FROM EMOJI ON TOURISM

Kanokkant Dansawangkul

A Project Submitted in Partial Fulfillment of the Requirements
for the Degree of Bachelor of Science Program in Computer Science

Department of Mathematics and Computer Science

Faculty of Science

Chulalongkorn University

Academic Year 2018

Copyright of Chulalongkorn University

หัวข้อโครงการ การวิเคราะห์ความรู้สึกจากอิมโม่จิเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์
โดย นางสาวกนกกานต์ ด้านสว่างกุล
สาขาวิชา วิทยาการคอมพิวเตอร์
อาจารย์ที่ปรึกษาโครงการ รองศาสตราจารย์ ดร.รัชติกา ลิปิกรณ์

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
อนุมัติให้นำโครงการฉบับนี้เป็นส่วนหนึ่ง ของการศึกษาตามหลักสูตรปริญญาบัณฑิต ในรายวิชา
2301499 โครงการวิทยาศาสตร์ (Senior Project)


.....
(ศาสตราจารย์ ดร. กฤษณะ เนียมมณี) หัวหน้าภาควิชาคณิตศาสตร์
และวิทยาการคอมพิวเตอร์

คณะกรรมการสอบโครงการ


.....
(รองศาสตราจารย์ ดร.รัชติกา ลิปิกรณ์) อาจารย์ที่ปรึกษาโครงการหลัก


.....
(ผู้ช่วยศาสตราจารย์ ดร.กรัง สีนอภิมย์สรานู) กรรมการ


.....
(รองศาสตราจารย์ ดร.วิมลรัตน์ งามอร่ามวางกูร) กรรมการ

นางสาวกนกกานต์ ด้านสว่างกุล: การวิเคราะห์ความรู้สึกจากอิโมจิเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์ (Twitter User's Sentiment Analysis from Emoji on Tourism). อ.ที่ปรึกษาโครงงานหลัก : รองศาสตราจารย์ ดร.รัชชิตา ลิปิกรณ์, 55 หน้า.

ในปัจจุบันผู้คนมีการใช้สื่อสังคมเพิ่มมากขึ้น โดยเฉพาะทวิตเตอร์ซึ่งเป็นสื่อสังคมอิสระที่ให้ผู้ใช้งานสามารถแสดงความคิดเห็น ประสบการณ์หรือความรู้สึกต่อสิ่งต่าง ๆ รวมไปถึงการเข้าถึงข้อมูลต่าง ๆ ก็สามารถทำได้ง่าย ทำให้ผู้ใช้งานสามารถนำข้อมูลเหล่านั้นมาช่วยในการตัดสินใจได้ด้วยตัวอย่างเช่นการนำข้อมูลบนทวิตเตอร์มาช่วยในการเลือกสถานที่ท่องเที่ยวต่าง ๆ ซึ่งส่งผลต่อเศรษฐกิจของสถานที่ท่องเที่ยวนั้น ๆ การวิเคราะห์ความรู้สึกของผู้สื่อสารผ่านสื่อสังคม จะสามารถนำมาใช้เพื่อประกอบการตัดสินใจเชิงธุรกิจในการพัฒนาและปรับปรุงสถานที่ท่องเที่ยวนั้นต่อไปได้ ซึ่งข้อมูลที่อยู่บนทวิตเตอร์จะมีทั้งข้อความ อิโมจิ และรูปภาพ ผู้จัดทำได้เห็นถึงความสำคัญของอิโมจิที่มีการใช้เพิ่มขึ้นเรื่อย ๆ และมีความสำคัญอย่างยิ่งต่อการวิเคราะห์ว่าผู้ที่แสดงความคิดเห็นนั้นมีความรู้สึกอย่างไร แต่ก็ยังมีผู้นำอิโมจิมาวิเคราะห์ไม่มากนัก ผู้จัดทำจึงได้พัฒนาระบบวิเคราะห์ความรู้สึกจากอิโมจิเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์ขึ้น โดยมีวัตถุประสงค์เพื่อวิเคราะห์ความรู้สึกของผู้ใช้งานทวิตเตอร์ออกเป็น 3 แบบ ได้แก่ ความรู้สึกเชิงบวก ความรู้สึกเชิงลบ และความรู้สึกเป็นกลาง ขั้นตอนการพัฒนาประกอบด้วย 4 ขั้นตอน คือ การรวบรวมข้อความทวิตจากทวิตเตอร์ การเตรียมข้อมูลสำหรับการนำไปวิเคราะห์ การเตรียมชุดข้อมูลสำหรับการนำไปวิเคราะห์ และการวิเคราะห์ความรู้สึกจากอิโมจิ จากนั้นผู้จัดทำจะเก็บผลการวิเคราะห์ไว้ในฐานข้อมูลเพื่อให้สามารถนำไปใช้วิเคราะห์และจำแนกข้อความทวิตรวมกับข้อมูลประเภทอื่น ๆ และนำมาใช้ในการปรับปรุงและพัฒนาแหล่งท่องเที่ยวในประเทศไทยต่อไป


ภาควิชา...คณิตศาสตร์และวิทยาการคอมพิวเตอร์.....ลายมือชื่อนิสิต กนกกานต์ ด้านสว่างกุล
 สาขาวิชา...วิทยาการคอมพิวเตอร์.....ลายมือชื่อ อ.ที่ปรึกษาโครงงานหลัก รัชชิตา ลิปิกรณ์
 ปีการศึกษา...2561.....

5833603023: MAJOR COMPUTER SCIENCE

KEYWORDS : SENTIMENT ANALYSIS / EMOJI / TWITTER / TOURISM

KANOKKANT DANSAWANGKUL: TWITTER USER'S SENTIMENT ANALYSIS FROM
EMOJI ON TOURISM. ADVISOR : ASSOC. PROF. RAJALIDA LIPIKORN, Ph.D., 55 pp.

Nowadays, the number of people using social media increases tremendously. Twitter is one of several social media that allows users to express opinions, experiences or feelings towards various things including easy access to various information. Users can easily use this information to help them make decisions. For example, users can use information gained from Twitter to help them choose the tourist places of attraction which can affect the economy of those places. Sentiment analysis of those people communicating through social media can be used to make business decisions in the development and improvement of those places. Since the information on Twitter is in the form of text, emoji, and images, we have seen that emoji usage is getting more and more popular and emoji has become more and more important in analyzing the feelings of those who express their opinions on Twitter, but not that many sentiment analyses have been performed on emoji. Therefore, we developed a system to analyze the sentiment of Twitter users from emoji. The objective of the proposed sentiment analysis system based on emoji is to classify the sentiment of Twitter users into 3 categories which are positive, negative, and neutral. The system development process consists of 4 steps: collecting tweets, data preprocessing, preparing a data set for processing, and analyzing sentiment from emoji. Then, the results of the analysis are recorded in the database so that they can be used in combination with other forms of information on Twitter to further analyze and classify tweets on tourism in order to improve and develop the tourism industry in Thailand.

Department : Mathematics and Computer Science Student's Signature 

Field of Study : Computer Science Advisor's Signature 

Academic Year : 2018

กิตติกรรมประกาศ

การจัดทำโครงการวิเคราะห์ความรู้สึกจากอิมojisเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์ สามารถสำเร็จลุล่วงไปได้ด้วยดี ทั้งนี้เนื่องจากได้รับความอนุเคราะห์และความช่วยเหลือจากบุคลากรต่าง ๆ หลายท่าน ดังนี้

ขอขอบพระคุณรองศาสตราจารย์ ดร.รัชชิตา ลิปิกรณ์ อาจารย์ที่ปรึกษาโครงการที่คอยให้คำแนะนำ คำปรึกษาและข้อเสนอแนะทางวิชาการ อีกทั้งยังช่วยแก้ไขและชี้แนะตลอดการทำงานทั้งโครงการ

ขอขอบพระคุณคณะกรรมการสอบ ได้แก่ ผู้ช่วยศาสตราจารย์ ดร.กรุง สินอภิรมย์สรานู และรองศาสตราจารย์ ดร.วิมลรัตน์ งามอร่ามวรางกูร ที่ช่วยให้คำแนะนำและข้อเสนอแนะ ซึ่งทำให้เกิดแนวทางในการพัฒนาโครงการนี้ ให้มีความถูกต้องสมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณคุณพ่อและคุณแม่ที่คอยช่วยสนับสนุน ให้กำลังใจมาโดยตลอด

ขอขอบคุณเพื่อน ๆ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ สาขาวิชาวิทยาการคอมพิวเตอร์รุ่น 23 ที่คอยช่วยเหลือและให้คำแนะนำ จนทำให้ผ่านอุปสรรคต่าง ๆ ไปได้ด้วยดี

ท้ายที่สุดนี้ ขอขอบพระคุณทุกความกรุณาจากทุกท่านที่กล่าวมา รวมถึงบุคคลที่ไม่ได้กล่าวถึงไว้ ณ ที่นี้อีกครั้งหนึ่ง สำหรับความช่วยเหลือและคำแนะนำต่าง ๆ ซึ่งทำให้โครงการนี้ประสบผลสำเร็จลุล่วงไปด้วยดี

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญ	ช
สารบัญตาราง	ฅ
สารบัญภาพ	ญ
บทที่ 1 บทนำ	11
1.1 ความเป็นมาและเหตุผลการวิจัย	11
1.2 วัตถุประสงค์ของการวิจัย	12
1.3 ขอบเขตการวิจัย	12
1.4 ขั้นตอนการวิจัย	12
1.5 ประโยชน์ที่ได้รับ	13
1.6 โครงสร้างของรายงาน	14
บทที่ 2 ความรู้เบื้องต้นและงานวิจัยที่เกี่ยวข้อง	15
2.1 ความรู้เบื้องต้นเกี่ยวกับอีโมจิ	15
2.2 ความรู้สึกของอีโมจิ (Sentiment of Emojis)	15
2.3 ภาษาไพธอน (Python)	18
2.4 MySQL	19
2.5 คอนฟิวชันเมทริกซ์ (confusion matrix)	19
2.6 Spring XD	20
บทที่ 3 การวิเคราะห์และออกแบบระบบ	21
3.1 การรวบรวมข้อมูล	21

3.2	การเตรียมข้อมูลสำหรับการนำไปวิเคราะห์.....	22
3.3	การเตรียมชุดข้อมูลสำหรับการวิเคราะห์.....	23
3.4	การวิเคราะห์ความรู้สึจากอีโมจิ	23
บทที่ 4	ผลการทดสอบ.....	26
4.1	ผลการทดสอบ	26
4.2	การพัฒนาระบบ	30
บทที่ 5	ข้อสรุป	35
5.1	ข้อสรุป.....	35
5.2	ปัญหาและอุปสรรค.....	35
5.3	วิธีการแก้ปัญหา	35
	รายการอ้างอิง.....	36
	ภาคผนวก ก แบบเสนอหัวข้อโครงการ รายวิชา 2301399 Project Proposal ปีการศึกษา 2561 .	39
	ภาคผนวก ข อีโมจิที่นำมาใช้ในการวิเคราะห์	43
	ประวัติผู้เขียน	55

สารบัญตาราง

	หน้า
ตารางที่ 1.1 แสดงขั้นตอนและเวลาการดำเนินงาน.....	13
ตารางที่ 2.1 คอนฟิวชันเมทริกซ์	19
ตารางที่ 3.1 การหาจำนวนเฉลี่ยที่ปรากฏอีโมจิทั้งหมดในข้อความทวิตแต่ละระดับความรู้สึก.....	25
ตารางที่ 4.1 คอนฟิวชันเมทริกซ์ผลลัพธ์การทดสอบการจำแนกข้อความทวิตที่มีหนึ่งอีโมจิ.....	26
ตารางที่ 4.2 ประสิทธิภาพการจำแนกข้อความทวิตที่มีหนึ่งอีโมจิ.....	27
ตารางที่ 4.3 คอนฟิวชันเมทริกซ์ผลลัพธ์การทดสอบการจำแนกข้อความทวิตที่มีมากกว่าอีโมจิ.....	27
ตารางที่ 4.4 ประสิทธิภาพการจำแนกข้อความทวิตที่มีมากกว่าอีโมจิ.....	28
ตารางที่ 4.5 คอนฟิวชันเมทริกซ์ผลลัพธ์การทดสอบการจำแนกข้อความทวิตด้วยอีโมจิ.....	29
ตารางที่ 4.6 ประสิทธิภาพการจำแนกข้อความทวิตด้วยอีโมจิ.....	29

สารบัญภาพ

	หน้า
ภาพที่ 2.1 การลำดับความรู้สึกจากอีโมจิ (Emoji Sentiment Ranking).....	16
ภาพที่ 2.2 การจับคู่ความรู้สึก (Sentiment Map).....	18
ภาพที่ 3.1 ภาพรวมของขั้นตอนในการวิเคราะห์ความรู้สึก จากอีโมจิเกี่ยวกับการท่องเที่ยวของผู้ใช้ ทวิตเตอร์.....	21
ภาพที่ 3.2 ข้อมูลข้อความทวิตที่ถูกรวบรวม	22
ภาพที่ 3.3 ข้อมูลในแต่ละข้อความทวิตที่ถูกรวบรวม	22
ภาพที่ 3.4 ชุดข้อมูลการจัดลำดับความรู้สึกจากอีโมจิ (Emoji Sentiment Ranking)	23
ภาพที่ 3.5 แถบแสดงช่วงของแต่ละระดับความรู้สึก	25
ภาพที่ 4.1 ผลลัพธ์ที่ได้จากการรวบรวมข้อความทวิต	30
ภาพที่ 4.2 ผลลัพธ์ที่ได้จากการดึงข้อความทวิต	30
ภาพที่ 4.3 ตัวอย่างข้อมูลในแต่ละบรรทัดของข้อมูลที่ถูกรวบรวม.....	31
ภาพที่ 4.4 ตัวอย่างซอร์สโค้ดในการตัดแบ่งข้อความทวิต.....	32
ภาพที่ 4.5 ผลลัพธ์จากการเตรียมข้อความทวิต	32
ภาพที่ 4.6 ข้อมูลการจัดลำดับความรู้สึกจากอีโมจิ.....	32
ภาพที่ 4.7 ชุดข้อมูลที่ถูกจัดเก็บไว้ในฐานข้อมูล MySQL	33
ภาพที่ 4.8 ตัวอย่างซอร์สโค้ดในการวิเคราะห์และจำแนกข้อความทวิต.....	34
ภาพที่ 4.9 ผลลัพธ์จากการวิเคราะห์และจำแนกข้อความทวิต	34

บทที่ 1

บทนำ

1.1 ความเป็นมาและเหตุผลการวิจัย

ในปัจจุบันผู้คนมีการใช้งานสื่อสังคม (social media) เพิ่มขึ้นเรื่อย ๆ เพื่อติดต่อสื่อสาร และแลกเปลี่ยนข้อมูลต่าง ๆ เช่น ความคิดเห็น ความรู้สึก หรือประสบการณ์ โดยจะแสดงในรูปแบบของข้อความ รูปภาพ วิดีโอ หรืออีโมจิ เป็นต้น และมีงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความรู้สึกจากข้อความเป็นจำนวนมาก แต่การสื่อสารบนสื่อสังคมด้วยข้อความเพียงอย่างเดียวอาจจะไม่เพียงพอที่จะวิเคราะห์ได้อย่างถูกต้องว่าผู้ใช้งานที่แสดงข้อความนั้นมีความรู้สึกอย่างไรกับข้อความนั้น หรือผู้ใช้งานที่อ่านข้อความนั้นอาจเข้าใจข้อความนั้นด้วยความรู้สึกที่ตรงกันข้ามกับผู้แสดงข้อความนั้นออกมา อีโมจิจะช่วยให้เข้าใจความรู้สึกของผู้ที่แสดงข้อความนั้นได้แม่นยำมากขึ้น เนื่องจากอีโมจิเป็นอักขรภาพที่สามารถสื่อสารได้ดีจากลักษณะของอีโมจิหรือสัญลักษณ์ที่คนส่วนใหญ่เข้าใจความหมายตรงกัน และในปัจจุบันมีอีโมจินมากถึง 2,823 อีโมจิที่อยู่ในมาตรฐานยูนิโคด [1] ที่ผู้ใช้สามารถเลือกใช้ได้อย่างหลากหลายและสามารถสื่อความรู้สึกได้ตรงความต้องการมากขึ้น อีโมจิจึงเป็นข้อมูลรูปแบบหนึ่งที่น่าสนใจที่จะนำมาศึกษา

ทวิตเตอร์เป็นสื่อสังคมที่นิยมใช้กันอย่างกว้างขวาง Omnicore Agency [2] ได้เปิดเผยจำนวนผู้ใช้ทวิตเตอร์ว่ามีประมาณ 100 ล้านบัญชีที่มีการใช้งานทุกวัน ทวิตเตอร์ยังเป็นแหล่งข้อมูลที่มีจำนวนมากมหาศาลและมีเนื้อหาหลากหลายทั้งข้อมูลทางด้านเทคโนโลยี การเมืองการปกครอง เศรษฐกิจ ความบันเทิง รวมถึงข้อมูลเกี่ยวกับการท่องเที่ยว ซึ่งเป็นข้อมูลที่สามารถส่งผลกระทบต่ออุตสาหกรรม การท่องเที่ยวในแต่ละประเทศได้ ทั้งนี้ ถ้าหากมีผู้ใช้ทวิตเตอร์แสดงความคิดเห็นต่อสถานที่ท่องเที่ยวในเชิงบวก จะช่วยส่งเสริมการท่องเที่ยวให้กับสถานที่ท่องเที่ยวนั้น และผู้ใช้ที่เห็นความคิดเห็นนี้อาจจะมีการแบ่งปันข้อมูลส่งต่อให้กับผู้ใช้คนอื่น ๆ ให้รู้จักกันต่อไปในวงกว้าง ในทางกลับกัน ถ้าหากมีผู้ใช้ทวิตเตอร์แสดงความคิดเห็นต่อสถานที่ท่องเที่ยวในเชิงลบ ผู้ที่เป็นเจ้าของสถานที่หรือผู้ที่ทำหน้าที่ในการรับผิดชอบต่อการดูแลสถานที่นั้น ๆ สามารถรับรู้ผลสะท้อนกลับ (feedback) จากการให้บริการของสถานที่ท่องเที่ยวนั้น และสืบหาวิธีการปรับปรุงและพัฒนาการให้บริการของสถานที่ท่องเที่ยวนั้น ๆ ได้จากความคิดเห็นเพิ่มเติมได้ จะช่วยให้เกิดประโยชน์ต่อสถานที่ท่องเที่ยวนั้น และยังเป็นการสร้างรายได้ให้กับประเทศ หรือบริเวณสถานที่ท่องเที่ยวนั้นด้วย

ผู้จัดทำจึงสนใจที่จะพัฒนาระบบวิเคราะห์ความรู้สึกของผู้ใช้ทวิตเตอร์จากอีโมจิที่เป็นรูปแบบข้อมูลที่ยังไม่ค่อยมีผู้ใดนำมาศึกษาและวิเคราะห์ ซึ่งจะนำข้อมูลเกี่ยวกับการท่องเที่ยวในทวิตเตอร์มาใช้เป็นกรณีศึกษา โดยจะนำ Emoji Sentiment Ranking [3] ที่ได้จากงานวิจัยของ P. Kralj Novak และคณะ [4] ซึ่งทำการวิเคราะห์ความรู้สึกและเสนอออกมาในรูปแบบของแถบความรู้สึก

(Sentiment bar) มาใช้เป็นแหล่งอ้างอิง และแบ่งความรู้สึกออกเป็น 3 แบบ ได้แก่ ความรู้สึกเชิงบวก (Positive) ความรู้สึกเชิงลบ (Negative) และความรู้สึกเป็นกลาง (Neutral)

1.2 วัตถุประสงค์ของการวิจัย

เพื่อพัฒนาระบบวิเคราะห์ความรู้สึกของผู้ใช้ทวิตเตอร์จากอิโมจิที่ผู้ใช้งานนั้นแสดงความคิดเห็นว่ามีความคิดเห็นเป็นบวก เป็นลบ หรือเป็นกลาง โดยใช้ข้อมูลเกี่ยวกับการท่องเที่ยวเป็นกรณีศึกษา เพื่อนำผลที่วิเคราะห์ได้มาใช้ในการปรับปรุงและพัฒนาแหล่งท่องเที่ยวให้ดียิ่งขึ้น

1.3 ขอบเขตการวิจัย

1. วิเคราะห์ความรู้สึกจากอิโมจิในแต่ละทวิตเท่านั้น
2. วิเคราะห์อิโมจิแค่ 751 อิโมจิเท่านั้น
3. ความรู้สึกของผู้ใช้งานจะถูกแบ่งเป็น 3 แบบ คือ เป็นบวก เป็นลบ หรือเป็นกลาง
4. ใช้ข้อมูลจากทวิตเตอร์เกี่ยวกับการท่องเที่ยวที่ประกอบด้วย #travel และอยู่ในปี พ.ศ. 2560 เป็นกรณีศึกษาสำหรับโครงการนี้เท่านั้น
5. ภาษาที่ใช้ในการสื่อสารผ่านทวิตเตอร์เป็นภาษาอังกฤษเท่านั้น
6. รหัสอิโมจิเป็นรหัสมาตรฐานบนระบบปฏิบัติการ Windows, IOS และ Android

1.4 ขั้นตอนการวิจัย

1. ศึกษาค้นหาข้อมูลที่เกี่ยวข้องกับการวิเคราะห์อิโมจิและความรู้สึก
2. ศึกษาเครื่องมือ โปรแกรมและเทคนิคที่ใช้ในโครงการ
3. กำหนดขอบเขตและวิธีการดำเนินงานของโครงการ
4. วิเคราะห์และออกแบบระบบวิเคราะห์ความรู้สึกของผู้ใช้ทวิตเตอร์จากอิโมจิ
5. พัฒนาระบบวิเคราะห์ความรู้สึกของผู้ใช้ทวิตเตอร์จากอิโมจิ
6. ทดสอบระบบวิเคราะห์ความรู้สึกของผู้ใช้ทวิตเตอร์จากอิโมจิ
7. ปรับปรุงและแก้ไขระบบวิเคราะห์ความรู้สึกของผู้ใช้ทวิตเตอร์จากอิโมจิ
8. สรุปผลการดำเนินการ
9. จัดทำเอกสาร

ตารางที่ 1.1 แสดงขั้นตอนและเวลาการดำเนินงาน

ขั้นตอนการดำเนินงาน	พ.ศ. 2561					พ.ศ. 2562			
	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.
1. ศึกษาค้นคว้าข้อมูลที่เกี่ยวข้อง									
2. ศึกษาเครื่องมือ โปรแกรมและเทคนิคที่ใช้									
3. กำหนดขอบเขตและวิธีการดำเนินงานของโครงการ									
4. วิเคราะห์และออกแบบระบบ									
5. พัฒนาระบบ									
6. ทดสอบระบบ									
7. ปรับปรุงและแก้ไขระบบ									
8. สรุปผลการดำเนินการ									
9. จัดทำเอกสาร									

1.5 ประโยชน์ที่ได้รับ

ก. ด้านความรู้และประสบการณ์ต่อตัวนิสิต

1. ได้ความรู้และความเข้าใจเกี่ยวกับวิธีการวิเคราะห์ความรู้สึกจากอีโมจิ
2. ได้เรียนรู้การใช้โปรแกรมที่นำมาใช้ในการพัฒนาซอฟต์แวร์
3. ได้เรียนรู้การวางแผนการทำงาน

ข. ประโยชน์ที่ได้จากโครงการที่พัฒนาขึ้น

1. ทำให้ทราบถึงความรู้สึกของผู้ใช้งานทวิตเตอร์ที่แสดงความคิดเห็นด้วยอีโมจิ
2. สามารถนำไปใช้ประกอบการตัดสินใจในการพัฒนาสถานที่ท่องเที่ยว
3. สามารถนำไปใช้ในการพัฒนาระบบอื่น ๆ ที่เกี่ยวกับการวิเคราะห์ความรู้สึกต่อไปได้

1.6 โครงสร้างของรายงาน

บทที่ 2 จะกล่าวถึงความรู้เบื้องต้นและงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความรู้สึกจากอิมโม่จิเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์

บทที่ 3 จะกล่าวถึงวิธีการวิจัย ซึ่งจะประกอบไปด้วยวิธีการรวบรวมข้อความทวิตจากทวิตเตอร์ การเตรียมข้อมูลสำหรับการนำไปวิเคราะห์ การเตรียมชุดข้อมูลสำหรับการวิเคราะห์ และการวิเคราะห์ความรู้สึกจากอิมโม่จิ

บทที่ 4 จะกล่าวถึงผลการทดสอบการวิเคราะห์ความรู้สึกจากอิมโม่จิเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์

บทที่ 5 จะกล่าวถึงข้อสรุปของการวิเคราะห์ความรู้สึกจากอิมโม่จิเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์

บทที่ 2

ความรู้เบื้องต้นและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงความรู้เบื้องต้นและงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความรู้สึกจากอิโมจิเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์ ดังต่อไปนี้

2.1 ความรู้เบื้องต้นเกี่ยวกับอิโมจิ

อิโมจิ [5] เป็นรูปภาพดิจิทัลที่ถูกใช้ในข้อความอิเล็กทรอนิกส์เพื่อแสดงความคิดหรือความรู้สึกเฉพาะ อิโมจินั้นมีหลากหลายประเภท เช่น การแสดงออกทางสีหน้า วัตถุ สถานที่ สัตว์ หรือสภาพอากาศ จะมีความคล้ายคลึงกับอิโมติคอน [6] แต่แตกต่างกันที่อิโมติคอนเป็นการพิมพ์ตัวอักษรที่แสดงเป็นใบหน้า เช่น :-) ซึ่งถูกคิดค้นขึ้นในยุคกำเนิดอินเทอร์เน็ตประมาณเดือนกันยายน ค.ศ. 1982 เมื่อนักวิทยาการคอมพิวเตอร์ Scott Fahlman แนะนำคณะกรรมการข้อความมหาวิทยาลัย Carnegie Mellon ว่า :-) และ :-(สามารถนำมาใช้เพื่อแยกแยะมุขตลกจากคำพูดที่ร้ายแรงทางออนไลน์ได้ หลังจากนั้นไม่นานสัญลักษณ์นั้นก็มีการใช้ชื่อว่า emotion icon ตรงกันข้ามกับอิโมจิที่เป็นรูปภาพจริงไม่ได้เกิดจากการพิมพ์ตัวอักษร เกิดขึ้นใน ค.ศ. 1990 โดย Shigetaka Kurita ซึ่งเป็นส่วนหนึ่งของทีมที่ทำงานใน NTT DoCoMo บริษัทสื่อสารของญี่ปุ่น ได้สร้างอิโมจิชุดแรกที่มีทั้งหมด 180 อิโมจิ โดยสร้างขึ้นจากการแสดงออกของคนในเมืองและสิ่งต่าง ๆ ที่เขาสังเกตเห็น [7]

อิโมจินั้นได้ถูกฝังอยู่ในบทสนทนาในชีวิตประจำวันของเรา ตามที่ Swyft Media [8] บอกว่า “ร้อยละ 74 ของผู้คนในสหรัฐที่ใช้สติ๊กเกอร์ [9] อิโมติคอนหรืออิโมจิในการสื่อสารออนไลน์ และส่งอิโมจิโดยเฉลี่ย 96 อิโมจิต่อวัน” เหตุผลสำหรับการใช้อิโมจิที่เพิ่มขึ้นเป็นอย่างมากก็คือลักษณะที่สร้างสรรค์เหล่านี้ ช่วยแสดงโทนเสียง ความรู้สึกและเพิ่มอารมณ์ขันให้กับบทสนทนาแบบดิจิทัล ซึ่งบางครั้งบริบทอาจหายไปเนื่องจากสื่อกลาง [10]

2.2 ความรู้สึกของอิโมจิ (Sentiment of Emojis)

งานวิจัย [4] เป็นงานวิจัยที่วิเคราะห์คุณสมบัติความรู้สึกของอิโมจิ โดยนำอิโมจิที่ถูกใช้บ่อยที่สุด 751 อิโมจิจากทวิตจำนวนประมาณ 1.6 ล้านทวิตของเจ้าของภาษา 18 ภาษาจำนวน 83 คน โดยระดับความรู้สึก ได้แก่ ความรู้สึกเชิงลบ ความรู้สึกเชิงบวก และความรู้สึกเป็นกลาง ซึ่งมีอิโมจิประกอบอยู่ 4% ของจำนวนทวิตทั้งหมดเหล่านี้ มาจัดทำคลังศัพท์ความรู้สึกของอิโมจิที่เรียกว่า การลำดับความรู้สึกจากอิโมจิ (Emoji Sentiment Ranking) และสร้างการจับคู่ความรู้สึก (Sentiment Map) ซึ่งระดับความรู้สึกของอิโมจินั้นคำนวณจากระดับความรู้สึกของทวิตที่อิโมจินั้นปรากฏ

การลำดับความรู้สึกจากอิโมจิ (Emoji Sentiment Ranking)

Char	Image [twemoji]	Unicode codepoint	Occurrences [5...max]	Position [0...1]	Neg [0...1]	Neu [0...1]	Pos [0...1]	Sentiment score [-1...+1]	Sentiment bar (c.i. 95%)	Unicode name	Unicode block
😊		0x1f802	14622	0.665	0.247	0.265	0.485	0.221		FACE WITH TEARS OF JOY	Emoticons
🖤		0x2764	6050	0.747	0.044	0.165	0.790	0.746		HEAVY BLACK HEART	Dingbats
🖤		0x2665	7144	0.754	0.035	0.272	0.693	0.657		BLACK HEART SUIT	Miscellaneous Symbols
😊		0x1f80d	6350	0.785	0.052	0.219	0.729	0.678		SMILING FACE WITH HEART-SHAPED EYES	Emoticons
😭		0x1f82d	5526	0.803	0.436	0.220	0.343	-0.093		LOUDLY CRYING FACE	Emoticons
😘		0x1f918	3648	0.654	0.053	0.193	0.754	0.701		FACE THROWING A KISS	Emoticons
😊		0x1f80a	3106	0.613	0.060	0.237	0.704	0.644		SMILING FACE WITH SMILING EYES	Emoticons
👌		0x1f44c	2925	0.805	0.094	0.249	0.657	0.563		OK HAND SIGN	Miscellaneous Symbols and Pictographs
💕		0x1f496	2400	0.766	0.042	0.205	0.674	0.632		TWO HEARTS	Miscellaneous Symbols and Pictographs
👏		0x1f44f	2336	0.787	0.104	0.271	0.624	0.520		CLAPPING HANDS SIGN	Miscellaneous Symbols and Pictographs

ภาพที่ 2.1 การลำดับความรู้สึกจากอิโมจิ (Emoji Sentiment Ranking)

จากภาพที่ 2.1 เป็นภาพการลำดับความรู้สึกจากอิโมจิ โดยแต่ละแถวจะแสดงข้อมูลของแต่ละอิโมจิ โดยมีข้อมูลดังนี้

Char	คืออิโมจิที่อยู่ในรูปของอักขระ
Image[twemoji]	คือภาพของอิโมจิที่ถูกแสดงบนเว็บไซต์
Unicode codepoint	คือตัวเลขที่เป็นรหัสมาตรฐานของแต่ละอิโมจิ
Occurrences[5...max]	คือจำนวนครั้งที่ปรากฏอิโมจิบนข้อความที่วัดทั้งหมดโดยมีค่าตั้งแต่ 5 ถึงค่าจำนวนครั้งที่ปรากฏมากที่สุด
Position[0...1]	คือตำแหน่งเฉลี่ยที่อิโมจินั้นถูกพบโดยมีค่าตั้งแต่ 0 ถึง 1 แทนการพบอิโมจิตั้งแต่ต้นข้อความถึงท้ายข้อความ
Neg[0...1]	คือค่าความน่าจะเป็นที่อิโมจิปรากฏอยู่บนข้อความเชิงลบ
Neu[0...1]	คือค่าความน่าจะเป็นที่อิโมจิปรากฏอยู่บนข้อความที่เป็นกลาง
Pos[0...1]	คือค่าความน่าจะเป็นที่อิโมจิปรากฏอยู่บนข้อความเชิงบวก
Sentiment score[-1...+1]	คือคะแนนความรู้สึกโดยมีค่าอยู่ในช่วง -1 ถึง 1
Sentiment bar (c.i. 95%)	คือแถบที่แสดงสัดส่วนของการปรากฏอิโมจิในแต่ละระดับความรู้สึก และคะแนนความรู้สึกในช่วงความเชื่อมั่น 95%
Unicode name	คือชื่อของอิโมจิ
Unicode block	คือกลุ่มของอิโมจิ

โดยความรู้สึกของแต่ละทวิตสามารถแบ่งออกเป็นเชิงบวก (positive) เป็นกลาง (neutral) และเชิงลบ (negative) [4] กำหนดให้ตัวแปร c แทนระดับความรู้สึกด้วยเลขจำนวนเต็ม 3 ค่า

$$c \in \{-1, 0, 1\}$$

โดยที่ -1 แทนความรู้สึกเชิงลบ 0 แทนความรู้สึกเป็นกลาง และ 1 แทนความรู้สึกเชิงบวก ในการวัดระดับความรู้สึกของอิโมจิแต่ละตัว และถ้านำจำนวนการปรากฏของความรู้สึกของอิโมจิมา รวมกันจะได้ดังสมการที่ (1)

$$\sum_c N(c) = N, c \in \{-1, 0, 1\} \quad (1)$$

โดยที่ N คือจำนวนการปรากฏของอิโมจิในทวิตทั้งหมด และ $N(c)$ คือจำนวนการปรากฏของอิโมจิในทวิตที่มีระดับความรู้สึก c และเมื่อนำมาสร้างการแจกแจงความน่าจะเป็นแบบไม่ต่อเนื่อง จะได้ดังสมการที่ (2)

$$\sum_c p_c = 1 \quad (2)$$

โดยที่ p_- , p_0 และ p_+ แสดงถึงความน่าจะเป็นที่อิโมจิปรากฏอยู่บนข้อความเชิงลบ ความน่าจะเป็นที่อิโมจิปรากฏอยู่บนข้อความที่เป็นกลาง และความน่าจะเป็นที่อิโมจิปรากฏอยู่บนข้อความเชิงบวก

โดยทั่วไปความน่าจะเป็นจะประมาณจากความถี่สัมพัทธ์ $p_c = N(c) / N$ ซึ่งเป็นการประมาณที่ดีสำหรับตัวอย่างขนาดใหญ่ แต่เนื่องจากกลุ่มตัวอย่างที่ได้มาส่วนใหญ่มีขนาดเล็ก ผู้จัดทำจึงใช้การประมาณของลาปลาซ (Laplace) ดังสมการที่ (3)

$$p_c = \frac{N(c)+1}{N+k} \quad (3)$$

เมื่อ c คือระดับความรู้สึก ได้แก่ ความรู้สึกเชิงลบ ความรู้สึกเป็นกลาง และความรู้สึกเชิงบวก และ k คือจำนวนของระดับความรู้สึก ซึ่งในที่นี้ $k = 3$

และสามารถคำนวณคะแนนความรู้สึก (sentiment score) โดยการแจกแจงความน่าจะเป็นแบบไม่ต่อเนื่อง (5) และคำนวณจากสมการ (4)

$$\bar{s} = -1 \cdot p_- + 0 \cdot p_0 + 1 \cdot p_+ = p_+ - p_- \quad (4)$$

โดยคะแนนความรู้สึกจะอยู่ในช่วง $-1 < \bar{s} < 1$

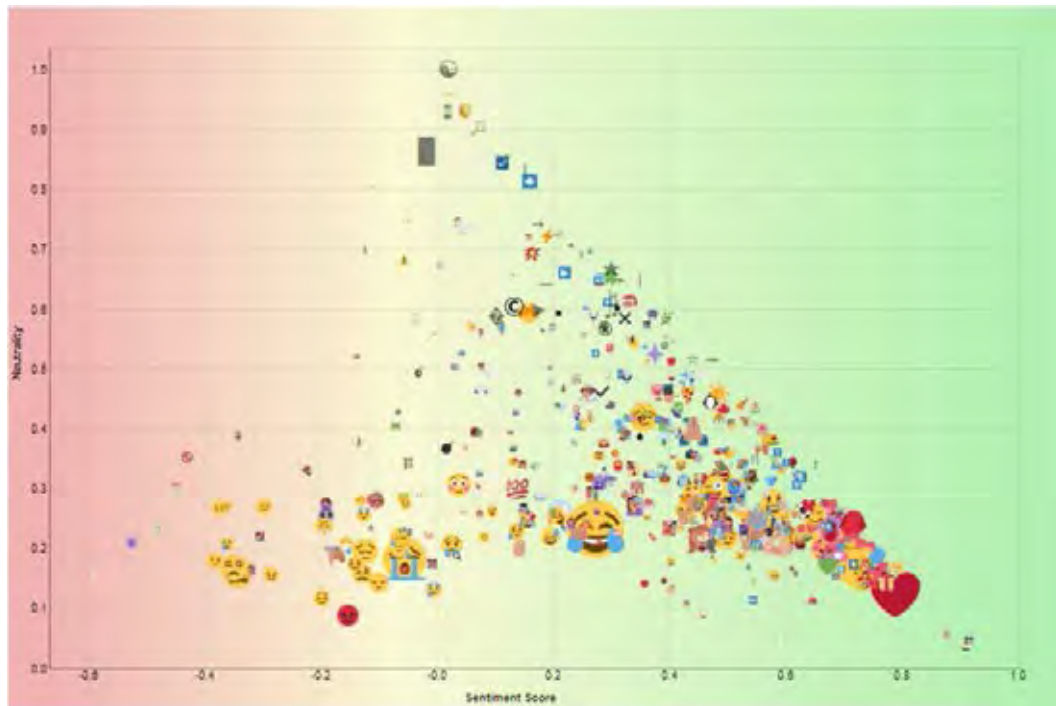
ค่าเบี่ยงเบนมาตรฐานของการแจกแจงความน่าจะเป็นแบบไม่ต่อเนื่อง (standard deviation: sd) คือ

$$sd = \sqrt{\sum_c p_c \cdot (c - \bar{s})^2} \quad (5)$$

และค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย (standard error of measurement: *sem*) คือ

$$sem = \frac{sd}{\sqrt{N}} \quad (6)$$

การจับคู่ความรู้สึก (Sentiment Map)



ภาพที่ 2.2 การจับคู่ความรู้สึก (Sentiment Map)

ภาพที่ 2.2 เป็นภาพการจับคู่ความรู้สึกโดยรวมของอีโมจิ 751 อีโมจิ ตำแหน่งของอีโมจิขึ้นอยู่กับคะแนนความรู้สึกและค่าความเป็นกลาง (p_0) ของอีโมจินั้น อีโมจิที่มีคะแนนความรู้สึกมากจะอยู่ทางด้านขวาหรือบริเวณสีเขียว อีโมจิที่มีคะแนนความรู้สึกน้อยจะอยู่ทางด้านซ้ายหรือบริเวณสีแดง และขนาดของอีโมจิขึ้นอยู่กับจำนวนครั้งที่ปรากฏอีโมจินั้นข้อความทวิตทั้งหมด

2.3 ภาษาไพธอน (Python)

Python [11] เป็นภาษาโปรแกรมที่เป็นที่นิยมภาษาหนึ่ง ถูกสร้างขึ้นโดย Guido Van Rossum และเปิดตัวในปี ค.ศ. 1991 ถูกใช้สำหรับการพัฒนาเว็บเซิร์ฟเวอร์ การพัฒนาซอฟต์แวร์ การคำนวณและการเขียนสคริปต์ระบบ ถูกออกแบบให้โปรแกรมสามารถถูกอ่านได้ง่ายขึ้น และโครงสร้างของภาษาทำให้โปรแกรมเมอร์สามารถเข้าใจแนวคิดการเขียนโปรแกรมได้ง่าย และใช้จำนวนบรรทัดในการเขียนโปรแกรมน้อยลงกว่าภาษา C++ และ Java โดยจะทำงานที่ละคำสั่ง มีการประมวลผลทันทีและทำงานในลักษณะที่สามารถโต้ตอบได้ ภาษาไพธอน [12] มีลักษณะเป็นระบบประเภทใด

นามิกและการจัดการหน่วยความจำอัตโนมัติ สนับสนุนกระบวนทัศน์การเขียนโปรแกรมทั้ง 4 กระบวนทัศน์หลัก ได้แก่ การเขียนโปรแกรมเชิงวัตถุ (object-oriented programming) การเขียนโปรแกรมเชิงคำสั่ง (imperative programming) การเขียนโปรแกรมเชิงฟังก์ชัน (functional programming) และการเขียนโปรแกรมเชิงตรรกะ (logic programming) นอกจากนี้ยังมีคลังโปรแกรม (library) มาตรฐานที่ครอบคลุม

2.4 MySQL

MySQL [13] เป็นระบบจัดการฐานข้อมูลเอสควิแอล (SQL) แบบโอเพ่นซอร์สที่ได้รับความนิยมสูงสุด ได้รับการพัฒนาแจกจ่ายและสนับสนุนโดย Oracle Corporation ซึ่งฐานข้อมูลคือชุดของข้อมูลที่มีโครงสร้าง ที่อาจเป็นอะไรก็ได้ตั้งแต่รายการช้อปปิ้งแบบง่ายไปจนถึงแกเลอรีรูปภาพหรือข้อมูลจำนวนมหาศาลในเครือข่ายองค์กร เพื่อเพิ่มข้อมูล เข้าถึงข้อมูลและประมวลผลข้อมูลที่เก็บไว้ในฐานข้อมูลในคอมพิวเตอร์ ซึ่งต้องมีระบบการจัดการฐานข้อมูลเช่น MySQL Server เนื่องจากคอมพิวเตอร์สามารถจัดการข้อมูลจำนวนมากได้ดี ระบบการจัดการฐานข้อมูลจึงมีบทบาทสำคัญในการใช้งานแบบเดี่ยว (stand-alone utility) หรือเป็นส่วนหนึ่งของแอปพลิเคชันอื่น ๆ

2.5 คอนฟิวชันเมทริกซ์ (confusion matrix)

คอนฟิวชันเมทริกซ์ [14] เป็นการประเมินผลลัพธ์การทำนายหรือผลลัพธ์จากโปรแกรมเปรียบเทียบกับผลลัพธ์จริงที่ถูกจำแนกโดยคน โดยมีรูปแบบเป็นตาราง 2 มิติระหว่างประเภทจริง (Actual class) กับประเภทที่ถูกทำนาย (Predicted class) ที่แสดงให้เห็นประสิทธิภาพการทำงานของโปรแกรม ดังตารางที่ 2.1

ตารางที่ 2.1 คอนฟิวชันเมทริกซ์

		Predicted Class	
		Yes	No
Actual Class	Yes	TP (true positive)	FN (false negative)
	No	FP (false positive)	TN (true negative)

โดยที่ TP คือจำนวนครั้งที่โปรแกรมทำนายว่าจริงและคนบอกว่าจริง TN คือจำนวนครั้งที่โปรแกรมทำนายว่าไม่จริงและคนบอกว่าไม่จริง FP คือจำนวนครั้งที่โปรแกรมทำนายว่าจริง แต่คนบอกว่าไม่จริง และ FN คือจำนวนครั้งที่โปรแกรมทำนายว่าไม่จริง แต่คนบอกว่าจริง

ค่า TP, TN, FP และ FN จะถูกใช้ในการคำนวณเพื่อสร้างเป็นเกณฑ์ต่าง ๆ ในการวัดค่าความถูกต้องในการจำแนกประเภท ได้แก่

1. ค่าความถูกต้องโดยรวม (Accuracy) เป็นเกณฑ์การวัดความถูกต้องโดยรวมในการจำแนกประเภท เพื่อพิจารณาการทำนายของตัวแบบว่าถูกต้องมากน้อยเพียงใด โดยนับจำนวนที่ทำนายถูกต้องทั้งหมดในทุกประเภทเทียบกับจำนวนตัวอย่างทั้งหมด ดังสมการ (7)

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (7)$$

2. ความถูกต้องสำหรับประเภทที่สนใจ (Recall หรือ True positive rate) เป็นเกณฑ์การวัดความถูกต้องในการจำแนกประเภท โดยมีมุมมองในการพิจารณาว่าหากความจริงเป็นประเภทที่สนใจ ตัวแบบจะสามารถทำนายได้ถูกกี่เปอร์เซ็นต์ ดังสมการ (8)

$$True\ positive\ rate = \frac{TP}{(TP+FN)} \quad (8)$$

3. ค่าความแม่นยำในการทำนายสำหรับประเภทที่สนใจ (Precision) เป็นเกณฑ์การวัดความถูกต้องในการจำแนกประเภท โดยมีมุมมองในการพิจารณาว่าหากตัวแบบทำนายประเภทที่สนใจ แล้วมีกี่เปอร์เซ็นต์ที่ทำนายได้ถูกต้อง ดังสมการ (9)

$$Precision = \frac{TP}{(TP+FP)} \quad (9)$$

4. ค่าประสิทธิภาพโดยรวม (F-measure) เป็นเกณฑ์การวัดความถูกต้องในการจำแนกประเภทที่ใช้ค่าเฉลี่ยฮาร์โมนิระหว่าง Recall และ Precision คำนวณได้ดังสมการ (10)

$$F - measure = \frac{2\ Recall * Precision}{Recall + Precision} \quad (10)$$

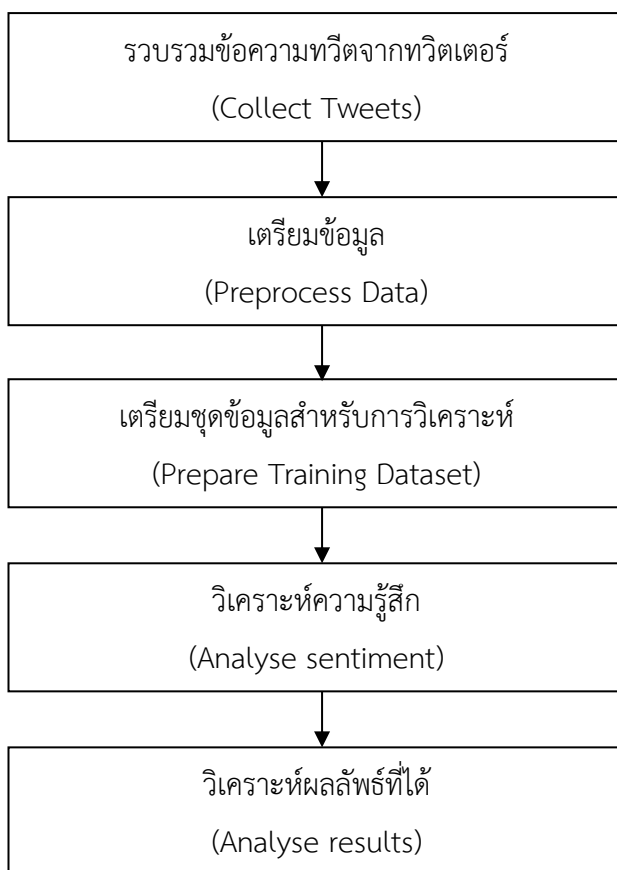
2.6 Spring XD

Spring XD [15] เป็นระบบแบบครบวงจร แบบกระจายและแบบขยายได้ สำหรับการนำเข้าข้อมูล การวิเคราะห์ตามเวลาจริง การประมวลผลแบบเชิงกลุ่มและการส่งออกข้อมูล โครงการ Spring XD เป็นโครงการโอเพ่นซอร์สลิขสิทธิ์ของ Apache 2 ที่มีเป้าหมายคือจัดการกับความซับซ้อนของข้อมูลขนาดใหญ่ ความซับซ้อนจำนวนมากในการสร้างแอปพลิเคชันข้อมูลขนาดใหญ่ในโลกแห่งความเป็นจริงนั้นเกี่ยวข้องกับการรวมระบบที่แตกต่างกันจำนวนมากเข้าไว้ในโซลูชันที่เชื่อมโยงกันเป็นหนึ่งเดียว กรณีใช้งานทั่วไปที่พบในการสร้างโซลูชันข้อมูลขนาดใหญ่ที่ครอบคลุมคือ การรับส่งข้อมูลปริมาณมากกระจายจากแหล่งนำเข้าที่หลากหลายไปยังแหล่งข้อมูลขนาดใหญ่เช่น HDFS หรือ Splunk การวิเคราะห์เวลาที่นำเข้าตามเวลาจริง เช่น การรวบรวมตัวชี้วัดและการนับค่า การจัดการการไหลของงาน (work flow) ผ่านงานเชิงกลุ่ม งานรวมการโต้ตอบกับระบบขององค์กรมาตรฐาน เช่น RDBMS รวมถึงการทำงานของ Hadoop เช่น MapReduce, HDFS, Pig, Hive หรือ HBase การส่งออกข้อมูลปริมาณงานมากเช่น จาก HDFS เป็นฐานข้อมูล RDBMS หรือ NoSQL

บทที่ 3

การวิเคราะห์และออกแบบระบบ

ในบทนี้จะกล่าวถึง วิธีการรวบรวมข้อความทวิตจากทวิตเตอร์ การเตรียมข้อมูลสำหรับการนำไปวิเคราะห์ การเตรียมชุดข้อมูลสำหรับการนำไปวิเคราะห์ และการวิเคราะห์ความรู้สึกจากอิโมจิ โดยภาพรวมของขั้นตอนในการวิเคราะห์ความรู้สึกจากอิโมจิเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์ สามารถสรุปได้เป็นดังภาพที่ 3.1



ภาพที่ 3.1 ภาพรวมของขั้นตอนในการวิเคราะห์ความรู้สึกจากอิโมจิเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์

3.1 การรวบรวมข้อมูล

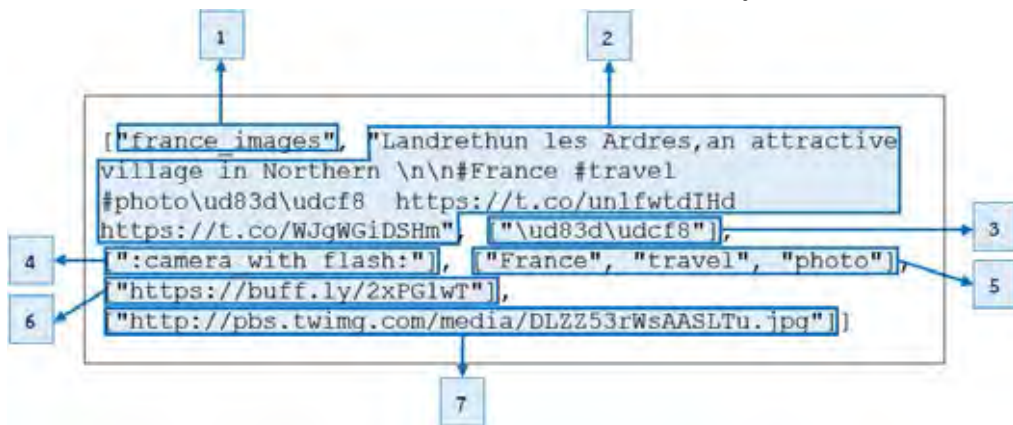
ผู้จัดทำได้รวบรวมข้อความทวิตที่จะนำมาใช้ในการวิเคราะห์จากทวิตเตอร์ เนื่องจากทวิตเตอร์เป็นสื่อสังคมที่มีข้อมูลและใช้งานเป็นจำนวนมาก โดยจะเลือกข้อความความคิดเห็นที่เป็น

ภาษาอังกฤษด้านการท่องเที่ยวมาใช้ในการวิเคราะห์ซึ่งอยู่ในปี พ.ศ. 2560 และรวบรวมจาก #travel โดยใช้ Spring XD เก็บอยู่ในรูปของไฟล์ .txt ดังภาพที่ 3.2



ภาพที่ 3.2 ข้อมูลข้อความทวีตที่ถูกรวบรวม

ข้อมูลที่ได้แต่ละบรรทัดหมายถึงข้อความในแต่ละทวีต ประกอบไปด้วยข้อมูลดังภาพที่ 3.3 โดยหมายเลข 1 คือชื่อผู้ใช้ทวีตเตอร์ หมายเลข 2 คือข้อความทวีต หมายเลข 3 คือกลุ่มของอิโมจิ ทั้งหมดที่มีอยู่ในข้อความทวีต ซึ่งอยู่ในรูปของ Escape Unicode หมายเลข 4 คือกลุ่มของชื่ออิโมจิ ทั้งหมดที่มีอยู่ในข้อความทวีต หมายเลข 5 คือกลุ่มของแฮชแท็ก (Hashtag) ทั้งหมดที่มีอยู่ในข้อความทวีต หมายเลข 6 ลิงค์ (Link) ของเว็บไซต์ และหมายเลข 7 คือลิงค์ของรูปภาพ



ภาพที่ 3.3 ข้อมูลในแต่ละข้อความทวีตที่ถูกรวบรวม

3.2 การเตรียมข้อมูลสำหรับการนำไปวิเคราะห์

เนื่องจากผู้จัดทำสนใจที่จะวิเคราะห์ความรู้สึกจากอิโมจิ ผู้จัดทำจึงนำแค่อิโมจิในแต่ละข้อความทวีตมาวิเคราะห์เท่านั้น ข้อความที่เหลือจะนำไปพิจารณาในโครงการอื่น ๆ ต่อไป โดยจะนำอิ

โมจิในแต่ละข้อความทวิตที่รวบรวมมาซึ่งอยู่ในรูปของ Escape Unicode มาแปลงเป็น Unicode code point ซึ่งเป็นรูปแบบตัวเลขฐานสิบหก เพื่อนำไปใช้เทียบกับข้อมูลในชุดข้อมูลที่เตรียมไว้

3.3 การเตรียมชุดข้อมูลสำหรับการวิเคราะห์

ผู้จัดทำได้นำข้อมูลการจัดลำดับความรู้สึกจากอิโมจิ (Emoji Sentiment Ranking) จากงานวิจัย [4] มาใช้ในการสร้างแบบจำลอง โดยเก็บข้อมูลไว้ในฐานข้อมูล MySQL ดังภาพที่ 3.2

emoji	unicode_code_point	occurrences	position	negative	neutral	positive	unicode_name	unicode_block
😭	0x1f602	14622	0.805101	3614	4163	6845	FACE WITH TEARS OF JOY	Emoticons
❤️	0x2764	8050	0.746943	355	1334	6361	HEAVY BLACK HEART	Dingbats
♥	0x2665	7144	0.753806	252	1942	4950	BLACK HEART SUIT	Miscellaneous Symbols
😊	0x1f60d	6359	0.765292	329	1390	4640	SMILING FACE WITH HEART-SHAPED EYES	Emoticons
😭	0x1f62d	5526	0.803352	2412	1218	1896	LOUDLY CRYING FACE	Emoticons
😘	0x1f618	3648	0.85448	193	702	2753	FACE THROWING A KISS	Emoticons
😊	0x1f60a	3186	0.813302	189	754	2243	SMILING FACE WITH SMILING EYES	Emoticons
👌	0x1f44c	2925	0.805223	274	728	1923	OK HAND SIGN	Miscellaneous Symbols and Pictographs
❤️	0x1f495	2400	0.765726	99	683	1618	TWO HEARTS	Miscellaneous Symbols and Pictographs
👏	0x1f44f	2336	0.78713	243	634	1459	CLAPPING HANDS SIGN	Miscellaneous Symbols and Pictographs
😊	0x1f601	2189	0.796151	278	648	1263	GRINNING FACE WITH SMILING EYES	Emoticons
😊	0x263a	2062	0.798634	128	449	1485	WHITE SMILING FACE	Miscellaneous Symbols
☺	0x2661	1975	0.763695	102	448	1425	WHITE HEART SUIT	Miscellaneous Symbols
👍	0x1f44d	1854	0.812126	213	460	1181	THUMBS UP SIGN	Miscellaneous Symbols and Pictographs
😓	0x1f629	1808	0.826214	1069	336	403	WEARY FACE	Emoticons
🙏	0x1f64f	1539	0.793848	124	648	767	PERSON WITH FOLDED HANDS	Emoticons
👊	0x270c	1534	0.79048	173	476	885	VICTORY HAND	Dingbats
😏	0x1f60f	1522	0.764977	170	676	676	SMIRKING FACE	Emoticons
😉	0x1f609	1521	0.844833	151	513	857	WINKING FACE	Emoticons

ภาพที่ 3.4 ชุดข้อมูลการจัดลำดับความรู้สึกจากอิโมจิ (Emoji Sentiment Ranking)

ผู้จัดทำได้เก็บข้อมูลอิโมจิลงในฐานข้อมูล โดยมีโครงสร้างของตารางที่ประกอบด้วย 9 คอลัมน์ ได้แก่


- emoji คืออิโมจิที่อยู่ในรูปของอักขระ
- unicode_code_point คือตัวเลขที่เป็นรหัสมาตรฐานของแต่ละอิโมจิ
- occurrences คือจำนวนครั้งที่ปรากฏอิโมจิบนข้อความทวิตทั้งหมด
- position คือตำแหน่งเฉลี่ยที่อิโมจินั้นถูกพบโดยมีค่าตั้งแต่ 0 ถึง 1
- negative แทนการพบอิโมจิตั้งแต่ต้นข้อความถึงท้ายข้อความ
- neutral คือจำนวนครั้งที่อิโมจิปรากฏอยู่บนข้อความเชิงลบ
- positive คือจำนวนครั้งที่อิโมจิปรากฏอยู่บนข้อความที่เป็นกลาง
- unicode_name คือชื่อของอิโมจิ
- unicode_block คือกลุ่มของอิโมจิ

3.4 การวิเคราะห์ความรู้สึกจากอิโมจิ

หลังจากที่เตรียมข้อมูลในขั้นตอนที่ 3.2 และ 3.3 เรียบร้อยแล้ว ผู้จัดทำจะวิเคราะห์อิโมจิในแต่ละข้อความทวิตโดยแบ่งออกเป็น 2 กรณี ดังนี้

กรณีที่ 1 ในข้อความทวิตนั้นมีเพียง 1 อีโมจิ

ผู้จัดทำจะนำ Unicode code point ที่ได้มาเปรียบเทียบกับข้อมูลที่จัดเก็บไว้ในฐานข้อมูล เพื่อดึงข้อมูลที่จะใช้ในการวิเคราะห์ความรู้สึกออกมา โดยนำค่าความน่าจะเป็นที่เป็นเชิงลบ เป็นกลาง และเชิงบวก มาทำให้อยู่ในช่วง -1 ถึง 1 และนำคะแนนความรู้สึกมาเปรียบเทียบกับค่าว่าอยู่ในช่วงที่เป็นเชิงลบ เป็นกลาง หรือเชิงบวก

ตัวอย่างเช่น Unicode code point ที่มีรหัส 0x2764 จะหมายถึงอีโมจิที่เป็นรูปหัวใจ  มีค่าความน่าจะเป็นที่ความรู้สึกเป็นเชิงลบเท่ากับ 0.044 ค่าความน่าจะเป็นที่ความรู้สึกเป็นกลางเท่ากับ 0.166 และค่าความน่าจะเป็นที่ความรู้สึกเป็นเชิงบวกเท่ากับ 0.790 เมื่อนำมาทำให้อยู่ในช่วง -1 ถึง 1 จะได้ช่วงที่เป็นลบเท่ากับ [-1,-0.912] ช่วงที่เป็นกลางเท่ากับ [-0.912,-0.58] และช่วงที่เป็นบวกเท่ากับ [-0.58,1] และค่าคะแนนความรู้สึกเท่ากับ 0.746 ซึ่งอยู่ในช่วงที่เป็นบวกทำให้สรุปได้ว่าอีโมจินี้มีระดับความรู้สึกเป็นเชิงบวก

กรณีที่ 2 ในข้อความทวิตนั้นมีมากกว่า 1 อีโมจิ

ผู้จัดทำจะนำ Unicode code point ทั้งหมดที่ได้มาเปรียบเทียบเพื่อดึงข้อมูลจากฐานข้อมูลที่ละอีโมจิ จากนั้นทำให้จำนวนการปรากฏของอีโมจิ (occurrences) มีจำนวนเท่ากัน โดยตัวคูณที่คูณแล้วทำให้จำนวนการปรากฏของอีโมจิเท่ากัน จะนำไปคูณจำนวนการปรากฏอีโมจิในข้อความเชิงลบ จำนวนการปรากฏอีโมจิในข้อความที่เป็นกลางและจำนวนการปรากฏอีโมจิในข้อความเชิงบวกด้วย และนำมาคำนวณหาค่าเฉลี่ยของจำนวนการปรากฏของอีโมจิในข้อความทวิตที่มีระดับความรู้สึกต่าง ๆ ดังสมการ (11) - (13)




$$N_- = \frac{\sum_{i=1}^m N_i(-1)}{m} \quad (11)$$

$$N_0 = \frac{\sum_{i=1}^m N_i(0)}{m} \quad (12)$$




$$N_+ = \frac{\sum_{i=1}^m N_i(1)}{m} \quad (13)$$

เมื่อ N_c คือจำนวนเฉลี่ยที่ปรากฏอีโมจิทั้งหมดในทวิตที่มีระดับความรู้สึก c $N_i(c)$ คือจำนวนที่ปรากฏอีโมจินั้นในทวิตทั้งหมดที่มีระดับความรู้สึก c และ m คือจำนวนอีโมจิที่มีอยู่ในทวิตนั้น

หลังจากนั้นก็หาค่าความน่าจะเป็นของแต่ละระดับความรู้สึกจากการประมาณของลาปลาซ แล้วนำมาทำให้อยู่ในช่วง -1 ถึง 1 และนำคะแนนความรู้สึกมาเปรียบเทียบกับค่าว่าอยู่ในช่วงของระดับความรู้สึกใด

ตัวอย่างเช่น Unicode code point ทั้งหมดที่ได้มาคือ 0x2764 0x1f60d และ 0x1f60d ซึ่งหมายถึง    เมื่อดึงข้อมูลและคำนวณหาจำนวนเฉลี่ยที่ปรากฏอิโมจิทั้งหมดในข้อความทวิตแต่ละระดับความรู้สึก จะได้ดังตารางที่ 3.1

ตารางที่ 3.1 การหาจำนวนเฉลี่ยที่ปรากฏอิโมจิทั้งหมดในข้อความทวิตแต่ละระดับความรู้สึก

	occurrence	negative	neutral	positive
	8050	355	1334	6361
	6359	329	1390	4640
	$6359 \times \frac{8050}{6359} = 8050$	$329 \times \frac{8050}{6359} = 416$	$1390 \times \frac{8050}{6359} = 1760$	$4640 \times \frac{8050}{6359} = 5874$
	6359	329	1390	4640
	$6359 \times \frac{8050}{6359} = 8050$	$329 \times \frac{8050}{6359} = 416$	$1390 \times \frac{8050}{6359} = 1760$	$4640 \times \frac{8050}{6359} = 5874$
		$N_- = 396$	$N_0 = 1618$	$N_+ = 6036$

หลังจากนั้นคำนวณหาค่าความน่าจะเป็นจากการประมาณของลาปลาซ จะได้ค่าความน่าจะเป็นของความรู้สึกที่เป็นเชิงลบเท่ากับ 0.049 ค่าความน่าจะเป็นที่ของความรู้สึกที่เป็นกลางเท่ากับ 0.201 และค่าความน่าจะเป็นของความรู้สึกที่เป็นเชิงบวกเท่ากับ 0.750 เมื่อนำมาทำให้อยู่ในช่วง -1 ถึง 1 จะได้ช่วงที่เป็นเชิงลบเท่ากับ [-1,-0.902] ช่วงที่เป็นกลางเท่ากับ [-0.902,-0.500] และช่วงที่เป็นเชิงบวกเท่ากับ [-0.500,1] และค่าคะแนนความรู้สึกเท่ากับ 0.701 ซึ่งอยู่ในช่วงที่เป็นบวกทำให้สรุปได้ว่าข้อความทวิตที่มีอิโมจิทั้งสามตัวนี้มีระดับความรู้สึกเป็นเชิงบวก

นอกจากนี้ ถ้าหากมีค่าคะแนนความรู้สึกอยู่ในช่วงที่เป็นกลาง แต่ความน่าจะเป็นของข้อความทวิตที่เป็นกลางมีค่าน้อยมาก ดังภาพที่ 3.5 แสดงว่าข้อความทวิตนั้นมีความน่าจะเป็นข้อความทวิตเชิงลบหรือข้อความทวิตเชิงบวกมากกว่า จากการสังเกตของผู้จัดทำ อิโมจิส่วนใหญ่ที่อยู่ในกรณีนี้ให้ความรู้สึกเชิงลบ ผู้จัดทำจึงกำหนดให้ข้อความทวิตที่อยู่ในกรณีนี้มีระดับความรู้สึกเป็นลบ



ภาพที่ 3.5 แถบแสดงช่วงของแต่ละระดับความรู้สึก

บทที่ 4

ผลการทดสอบ

ในบทนี้จะกล่าวถึงผลการทดสอบของการวิเคราะห์ความรู้สึกจากอิโมจิเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์ โดยสามารถทดสอบความน่าเชื่อถือของผลลัพธ์ได้โดยพิจารณาจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าการเรียกคืน (Recall) และค่าวัดประสิทธิภาพโดยรวม (F-measure) และกล่าวถึงส่วนของการพัฒนาระบบการวิเคราะห์ความรู้สึกจากอิโมจิเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์ ดังรายละเอียดต่อไปนี้

4.1 ผลการทดสอบ

การทดสอบการจำแนกรู้สึกด้วยอิโมจิ จำนวน 450 ข้อความ โดยแบ่งเป็นข้อความทวิตเชิงบวก 150 ข้อความ ข้อความทวิตที่เป็นกลาง 150 ข้อความ และข้อความแสดงความคิดเห็นเชิงลบ 150 ข้อความ ผู้จัดทำจะแบ่งผลการทดสอบออกเป็น 2 กรณี ดังนี้

กรณีที่ข้อความทวิตมีหนึ่งอิโมจิ

กรณีที่ข้อความทวิตมีหนึ่งอิโมจิจะประกอบด้วยข้อความทวิตเชิงบวก 66 ข้อความ ข้อความทวิตที่เป็นกลาง 77 ข้อความ และข้อความแสดงความคิดเห็นเชิงลบ 59 ข้อความ และพิจารณาค่าความถูกต้อง ค่าความแม่นยำ ค่าการเรียกคืน รวมทั้งค่าวัดประสิทธิภาพโดยรวมในการจำแนกข้อมูลว่ามีประสิทธิภาพมากน้อยเพียงใด ผลการทดสอบการจำแนกข้อความทวิตดังตารางที่ 4.1 และ 4.2

ตารางที่ 4.1 คอนฟิวชันเมทริกซ์ผลลัพธ์การทดสอบการจำแนกข้อความทวิตที่มีหนึ่งอิโมจิ

		Predicted Class			ไม่มีข้อมูล อิโมจิ
		Positive	Neutral	Negative	
Actual Class	Positive	48	12	2	4
	Neutral	30	30	11	6
	Negative	4	1	54	0

จากตารางที่ 4.1 แสดงผลลัพธ์ในการจำแนกข้อความทวิตด้วยอิโมจิระหว่างคลาสข้อความทวิตเชิงบวก เป็นกลาง และเชิงลบ ทั้งหมด 202 ข้อความ โดยการจำแนกมีจำนวนข้อความเชิงบวกและแบบจำลองจำแนกถูกต้องเท่ากับ 48 ข้อความ จำนวนข้อความเชิงบวกแต่แบบจำลองจำแนกว่าเป็นข้อความที่เป็นกลางเท่ากับ 12 ข้อความ จำนวนข้อความเชิงบวกแต่แบบจำลองจำแนกว่าเป็นข้อความที่เชิงลบเท่ากับ 2 ข้อความ จำนวนข้อความที่เป็นกลางและแบบจำลองจำแนกถูกต้องเท่ากับ

30 ข้อความ จำนวนข้อความที่เป็นกลางแต่แบบจำลองจำแนกว่าเป็นข้อความเชิงบวกเท่ากับ 30 ข้อความ จำนวนข้อความที่เป็นกลางแต่แบบจำลองจำแนกว่าเป็นข้อความเชิงลบเท่ากับ 11 ข้อความ อีกทั้งจำนวนข้อความเชิงลบและแบบจำลองจำแนกถูกต้องเท่ากับ 54 ข้อความ จำนวนข้อความเชิงลบแต่แบบจำลองจำแนกว่าเป็นข้อความเชิงบวกเท่ากับ 4 ข้อความ และจำนวนข้อความเชิงลบแต่แบบจำลองจำแนกว่าเป็นข้อความที่เป็นกลางเท่ากับ 1 ข้อความ นอกจากนี้ยังมีข้อความทวีตที่ไม่มีข้อมูลของอิโมจิในแบบจำลองจำนวน 10 ข้อความ

ตารางที่ 4.2 ประสิทธิภาพการจำแนกข้อความทวีตที่มีหนึ่งอิโมจิ

	Accuracy	Precision	Recall	F-measure
Positive	65.35%	58.54%	72.73%	64.86%
Neutral		69.77%	38.96%	50.00%
negative		80.60%	91.53%	85.71%

จากตารางที่ 4.2 ผลการประเมินประสิทธิภาพของระบบในการจำแนกข้อความทวีตด้วยอิโมจิมีค่าความถูกต้องเฉลี่ยเท่ากับ 65.35% ค่าความแม่นยำของการจำแนกข้อความเชิงบวก เชิงกลาง และเชิงลบเท่ากับ 58.54% 69.77% และ 80.60% ตามลำดับ ซึ่งค่าความแม่นยำเฉลี่ยจะเท่ากับ 69.63% ค่าการเรียกคืนของการจำแนกข้อความเชิงบวก เชิงกลาง และเชิงลบเท่ากับ 72.73% 38.96% และ 91.53% ตามลำดับ ซึ่งค่าการเรียกคืนเฉลี่ยจะเท่ากับ 67.74% และค่าวัดประสิทธิภาพโดยรวมของการจำแนกข้อความเชิงบวก เชิงกลาง และเชิงลบเท่ากับ 64.86% 50.00% และ 85.71% ตามลำดับ ซึ่งค่าวัดประสิทธิภาพโดยรวมเฉลี่ยจะเท่ากับ 66.86%

กรณีที่ข้อความทวีตมีมากกว่าหนึ่งอิโมจิ

กรณีที่ข้อความทวีตมีหนึ่งอิโมจิจะประกอบด้วยข้อความทวีตเชิงบวก 84 ข้อความ ข้อความทวีตที่เป็นกลาง 73 ข้อความ และข้อความแสดงความคิดเห็นเชิงลบ 91 ข้อความ และพิจารณาค่าความถูกต้อง ค่าความแม่นยำ ค่าการเรียกคืน รวมทั้งค่าวัดประสิทธิภาพโดยรวมในการจำแนกข้อมูลว่ามีประสิทธิภาพมากน้อยเพียงใด ผลการทดสอบการจำแนกข้อความทวีตดังตารางที่ 4.3 และ 4.4

ตารางที่ 4.3 คอนฟิวชันเมทริกซ์ผลลัพธ์การทดสอบการจำแนกข้อความทวีตที่มีมากกว่าอิโมจิ

		Predicted Class			ไม่มีข้อมูล อิโมจิ
		Positive	Neutral	Negative	
Actual Class	Positive	76	3	3	2
	Neutral	23	39	6	5
	Negative	14	21	56	0

จากตารางที่ 4.3 แสดงผลลัพธ์ในการจำแนกข้อความทวิตด้วยโอมิเจอร์ระหว่างคลาสข้อความ ทวิตเชิงบวก เป็นกลาง และเชิงลบ ทั้งหมด 248 ข้อความ โดยการจำแนกมีจำนวนข้อความเชิงบวก และแบบจำลองจำแนกถูกต้องเท่ากับ 76 ข้อความ จำนวนข้อความเชิงบวกแต่แบบจำลองจำแนกว่า เป็นข้อความที่เป็นกลางเท่ากับ 3 ข้อความ จำนวนข้อความเชิงบวกแต่แบบจำลองจำแนกว่าเป็น ข้อความที่เชิงลบเท่ากับ 3 ข้อความ จำนวนข้อความที่เป็นกลางและแบบจำลองจำแนกถูกต้องเท่ากับ 39 ข้อความ จำนวนข้อความที่เป็นกลางแต่แบบจำลองจำแนกว่าเป็นข้อความเชิงบวกเท่ากับ 23 ข้อความ จำนวนข้อความที่เป็นกลางแต่แบบจำลองจำแนกว่าเป็นข้อความเชิงลบเท่ากับ 6 ข้อความ อีกทั้งจำนวนข้อความเชิงลบและแบบจำลองจำแนกถูกต้องเท่ากับ 56 ข้อความ จำนวนข้อความเชิง ลบแต่แบบจำลองจำแนกว่าเป็นข้อความเชิงบวกเท่ากับ 14 ข้อความ และจำนวนข้อความเชิงลบแต่ แบบจำลองจำแนกว่าเป็นข้อความที่เป็นกลางเท่ากับ 21 ข้อความ นอกจากนี้ยังมีข้อความทวิตที่ไม่มี ข้อมูลของโอมิเจอร์ในแบบจำลองจำนวน 7 ข้อความ

ตารางที่ 4.4 ประสิทธิภาพการจำแนกข้อความทวิตที่มีมากกว่าโอมิเจอร์

	Accuracy	Precision	Recall	F-measure
Positive	68.95%	67.26%	90.48%	77.16%
Neutral		61.90%	53.42%	57.35%
negative		86.15%	61.54%	71.79%

จากตารางที่ 4.4 ผลการประเมินประสิทธิภาพของระบบในการจำแนกข้อความทวิตด้วยโอมิเจอร์ มีค่าความถูกต้องเฉลี่ยเท่ากับ 68.95% ค่าความแม่นยำของการจำแนกข้อความเชิงบวก เชิงกลาง และเชิงลบเท่ากับ 67.26% 61.90% และ 86.15% ตามลำดับ ซึ่งค่าความแม่นยำเฉลี่ยจะเท่ากับ 71.77% ค่าการเรียกคืนของการจำแนกข้อความเชิงบวก เชิงกลาง และเชิงลบเท่ากับ 90.48% 53.42% และ 61.54% ตามลำดับ ซึ่งค่าการเรียกคืนเฉลี่ยจะเท่ากับ 68.48% และค่าวัดประสิทธิภาพ โดยรวมของการจำแนกข้อความเชิงบวก เชิงกลาง และเชิงลบเท่ากับ 77.16% 57.35% และ 71.79% ตามลำดับ ซึ่งค่าวัดประสิทธิภาพโดยรวมเฉลี่ยจะเท่ากับ 68.77%

ผลการทดสอบรวมทั้งสองกรณี

เมื่อรวมทั้งสองกรณีจะประกอบด้วยข้อความทวิตเชิงบวก 150 ข้อความ ข้อความทวิตที่เป็น กลาง 150 ข้อความ และข้อความแสดงความคิดเห็นเชิงลบ 150 ข้อความ และพิจารณาค่าความ ถูกต้อง ค่าความแม่นยำ ค่าการเรียกคืน รวมทั้งค่าวัดประสิทธิภาพโดยรวมในการจำแนกข้อมูลว่ามี ประสิทธิภาพมากน้อยเพียงใด ผลการทดสอบการจำแนกข้อความทวิตดังตารางที่ 4.5 และ 4.6

ตารางที่ 4.5 คอนฟิวชันเมทริกซ์ผลลัพธ์การทดสอบการจำแนกข้อความทวิตด้วยอิโมจิ

		Predicted Class			ไม่มีข้อมูล อิโมจิ
		Positive	Neutral	Negative	
Actual Class	Positive	124	16	4	6
	Neutral	53	69	17	11
	Negative	18	22	110	0

จากตารางที่ 4.5 แสดงผลลัพธ์ในการจำแนกข้อความทวิตด้วยอิโมจิระหว่างคลาสข้อความทวิตเชิงบวก เป็นกลาง และเชิงลบ ทั้งหมด 450 ข้อความ โดยการจำแนกมีจำนวนข้อความเชิงบวกและแบบจำลองจำแนกถูกต้องเท่ากับ 124 ข้อความ จำนวนข้อความเชิงบวกแต่แบบจำลองจำแนกว่าเป็นข้อความที่เป็นกลางเท่ากับ 16 ข้อความ จำนวนข้อความเชิงบวกแต่แบบจำลองจำแนกว่าเป็นข้อความที่เชิงลบเท่ากับ 4 ข้อความ จำนวนข้อความที่เป็นกลางและแบบจำลองจำแนกถูกต้องเท่ากับ 69 ข้อความ จำนวนข้อความที่เป็นกลางแต่แบบจำลองจำแนกว่าเป็นข้อความเชิงบวกเท่ากับ 53 ข้อความ จำนวนข้อความที่เป็นกลางแต่แบบจำลองจำแนกว่าเป็นข้อความเชิงลบเท่ากับ 17 ข้อความ อีกทั้งจำนวนข้อความเชิงลบและแบบจำลองจำแนกถูกต้องเท่ากับ 110 ข้อความ จำนวนข้อความเชิงลบแต่แบบจำลองจำแนกว่าเป็นข้อความที่เป็นกลางเท่ากับ 22 ข้อความ นอกจากนี้ยังมีข้อความทวิตที่ไม่มีข้อมูลของอิโมจิในแบบจำลองจำนวน 17 ข้อความ

ตารางที่ 4.6 ประสิทธิภาพการจำแนกข้อความทวิตด้วยอิโมจิ

	Accuracy	Precision	Recall	F-measure
Positive	67.33%	63.59%	82.67%	71.88%
Neutral		64.49%	46.00%	53.70%
negative		83.97%	73.33%	78.29%

จากตารางที่ 4.6 ผลการประเมินประสิทธิภาพของระบบในการจำแนกข้อความทวิตด้วยอิโมจิ มีค่าความถูกต้องเฉลี่ยเท่ากับ 67.33% ค่าความแม่นยำของการจำแนกข้อความเชิงบวก เชิงกลาง และเชิงลบเท่ากับ 63.59% 64.49% และ 83.97% ตามลำดับ ซึ่งค่าความแม่นยำเฉลี่ยจะเท่ากับ 70.68% ค่าการเรียกคืนของการจำแนกข้อความเชิงบวก เชิงกลาง และเชิงลบเท่ากับ 82.67% 46.00% และ 73.33% ตามลำดับ ซึ่งค่าการเรียกคืนเฉลี่ยจะเท่ากับ 67.33% และค่าวัดประสิทธิภาพโดยรวมของการจำแนกข้อความเชิงบวก เชิงกลาง และเชิงลบเท่ากับ 71.88% 53.70% และ 78.29% ตามลำดับ ซึ่งค่าวัดประสิทธิภาพโดยรวมเฉลี่ยจะเท่ากับ 67.96%

ผลสรุปการทดลอง

สังเกตได้ว่าผลการทดลองของกรณีที่ข้อความทวีตมีมากกว่าหนึ่งอิมojis จะดีกว่าผลการทดลองของกรณีที่ข้อความทวีตมีเพียงหนึ่งอิมoji ทั้งค่าความถูกต้อง ค่าความแม่นยำ ค่าการเรียกคืน และค่าวัดประสิทธิภาพโดยรวม เนื่องจากเมื่อข้อความทวีตมีหลายอิมojินั้นอาจมีความหมายว่าผู้ใช้ชีวิตเตอร์นั้นต้องการเน้นว่ามีความรู้สึกในระดับใดระดับหนึ่งเป็นอย่างมากจึงเพิ่มจำนวนอิมojiเข้าไปในข้อความทวีต ทำให้อิมojisมีน้ำหนักมากขึ้น ผลการจำแนกจึงถูกต้องมากขึ้น

4.2 การพัฒนาระบบ

ในหัวข้อนี้จะกล่าวถึงการพัฒนาาระบบ ซึ่งพัฒนาโดยใช้ภาษาไพธอน มีอยู่ 4 ขั้นตอน ได้แก่ การรวบรวมข้อความทวีตจากทวีตเตอร์ การเตรียมข้อความทวีตเพื่อนำไปวิเคราะห์ การเตรียมชุดข้อมูลสำหรับการวิเคราะห์ และการวิเคราะห์และจำแนกความรู้สึกจากอิมoji โดยมีรายละเอียดดังนี้

1. การรวบรวมข้อความทวีตจากทวีตเตอร์

ผู้จัดทำได้นำเครื่องมือ Spring XD ดังที่ได้กล่าวไว้ในบทที่ 2 มาใช้ในการรวบรวมข้อความทวีต จะได้ข้อมูลเป็นรูปแบบ JSON ดังภาพที่ 4.1 หลังจากนั้นจะเลือกเฉพาะข้อความที่อยู่ในปี พ.ศ. 2560 ซึ่งมีอิมojiอยู่ในข้อความทวีตและประกอบด้วย #travel ผลลัพธ์ที่ได้จากการดึงข้อความทวีตจะเป็นไฟล์ .txt ดังภาพที่ 4.2



ภาพที่ 4.1 ผลลัพธ์ที่ได้จากการรวบรวมข้อความทวีต



ภาพที่ 4.2 ผลลัพธ์ที่ได้จากการดึงข้อความทวีต

โดยข้อความแต่ละบรรทัดจะมีข้อมูลประเภทต่าง ๆ ได้แก่ ผู้ใช้ทวิตเตอร์ ข้อความทวิต กลุ่มของอีโมจิทั้งหมดที่มีอยู่ในข้อความทวิต กลุ่มของชื่ออีโมจิทั้งหมดที่มีอยู่ในข้อความทวิต กลุ่มของแฮชแท็กทั้งหมดที่มีอยู่ในข้อความทวิต ลิงค์ของเว็บไซต์ และลิงค์ของรูปภาพ แสดงดังภาพที่ 4.3

```
["tomonakata_com", "WELCOME TO THAILAND\nl-77 Province in
Thailand\nWelcome to visits\n#World #Thailand #Travel #Tour #Nature
#Bangkok\nhttps://t.co/XQSjETNimH
\u083c\u0df39\u083c\u0df39\u083c\u0df39 https://t.co/MSRkKrjTEU",
["\u083c\u0df39,\u083c\u0df39,\u083c\u0df39"], [":rose:,:rose:,:rose:"],
["World", "Thailand", "Travel", "Tour", "Nature", "Bangkok"], [], []]
```

ภาพที่ 4.3 ตัวอย่างข้อมูลในแต่ละบรรทัดของข้อมูลที่ถูกรวบรวม

2. การเตรียมข้อความทวิตเพื่อนำไปวิเคราะห์

หลังจากรวบรวมข้อมูลมาเป็นไฟล์ .txt แล้ว จะทำการตัดแบ่งข้อความออกตามประเภทต่าง ๆ ได้แก่

acc_tweet	เก็บข้อมูลผู้ใช้ทวิตเตอร์และข้อความทวิต
unicode_emoji	เก็บข้อมูล unicode ของอีโมจิทั้งหมดที่มีอยู่ในข้อความทวิต
unicode_name	เก็บข้อมูลชื่ออีโมจิทั้งหมดที่มีอยู่ในข้อความทวิต
hashtag	เก็บข้อมูลแฮชแท็กทั้งหมดที่มีอยู่ในข้อความทวิต
website	เก็บข้อมูลลิงค์ของเว็บไซต์
picture	เก็บข้อมูลลิงค์ของรูปภาพ

หลังจากนั้นผู้จัดทำจะนำ unicode_emoji มาแปลงจาก Escape Unicode เป็น Unicode code point ทีละข้อความทวิต แล้วเก็บไว้ในตัวแปร unicode ตัวอย่างการทำงานของโปรแกรม แสดงซอร์สโค้ดดังภาพที่ 4.4 ผลลัพธ์จากการเตรียมข้อความทวิตดังภาพที่ 4.5

```
import json
acc_tweet = []
unicode_emoji = []
unicode_name = []
hashtag = []
website = []
picture = []
with open('outfile_hashtag_emoji_only_noDup.txt') as f:
    while True:
        s = f.readline()
        if s == '':
            break
        c = s.replace('[]', '[""]')
        d = c.split('"');
        acc_tweet.append(d[1][0:-3])
        unicode_emoji.append(d[2][0:-4])
        unicode_name.append(d[3][0:-4])
        hashtag.append(d[4][0:-4])
        website.append(d[5][0:-4])
        picture.append(d[6][0:-4])
```



```

unicode = []
for row in range(len(unicode_emoji)):
    temp = str(unicode_emoji[row])
    emojis = temp.split(',')
    unicode_in_tweet = []
    for no in range(len(emojis)):
        emoji = '\"' + emojis[no] + '\"'
        code = ascii(json.loads(emoji))
        code = int(code[3:-1], 16)
        unicode_in_tweet.append(hex(code))
    unicode.append(unicode_in_tweet)

```

ภาพที่ 4.4 ตัวอย่างซอร์สโค้ดในการตัดแบ่งข้อความทวีต

```

['0x1f339', '0x1f339', '0x1f339'],
['0x1f483', '0x1f64c', '0x1f64c'],
['0x1f422'],
['0x1f609'],
['0x1f64c'],
['0x1f44d'],
...

```

ภาพที่ 4.5 ผลลัพธ์จากการเตรียมข้อความทวีต

3. การเตรียมชุดข้อมูลสำหรับการวิเคราะห์

ผู้จัดทำได้นำข้อมูลการจัดลำดับความรู้สึกจากอิโมจิ (Emoji Sentiment Ranking) จากงานวิจัย [4] ซึ่งอยู่ในรูปของไฟล์ .csv ดังภาพที่ 4.6 โดยในคอลัมน์ Emoji จะเป็นอิโมจิที่ถูกเข้ารหัสแบบ UTF-8 ซึ่งเป็นการเข้ารหัสชนิดหนึ่ง มาจัดเก็บไว้ในฐานข้อมูล MySQL ทั้งหมด โดยในคอลัมน์ emoji จะถอดรหัสกลับมาเป็นอิโมจิ ดังภาพที่ 4.7

	A	B	C	D	E	F	G	H	I
1	Emoji	Unicode codepoint	Occurrences	Position	Negative	Neutral	Positive	Unicode name	Unicode block
2	😭	0x1f602	14622	0.805101	3614	4163	6845	FACE WITH TEARS OF JOY	Emoticons
3	💩	0x2764	8050	0.746943	355	1334	6361	HEAVY BLACK HEART	Dingbats
4	🖤	0x2665	7144	0.753806	252	1942	4950	BLACK HEART SUIT	Miscellaneous Symbols
5	😊	0x1f60d	6359	0.765292	329	1390	4640	SMILING FACE WITH HEAR	Emoticons
6	😭	0x1f62d	5526	0.803352	2412	1218	1896	LOUDLY CRYING FACE	Emoticons
7	😘	0x1f618	3648	0.85448	193	702	2753	FACE THROWING A KISS	Emoticons
8	😊	0x1f60a	3186	0.813302	189	754	2243	SMILING FACE WITH SMILI	Emoticons
9	👌	0x1f44c	2925	0.805223	274	728	1923	OK HAND SIGN	Miscellaneous Symbols
10	💕	0x1f495	2400	0.765726	99	683	1618	TWO HEARTS	Miscellaneous Symbols
11	👏	0x1f44f	2336	0.78713	243	634	1459	CLAPPING HANDS SIGN	Miscellaneous Symbols
12	😊	0x1f601	2189	0.796151	278	648	1263	GRINNING FACE WITH SMI	Emoticons
13	👉	0x263a	2062	0.798634	128	449	1485	WHITE SMILING FACE	Miscellaneous Symbols
14	👉	0x2961	1975	0.783695	102	448	1425	WHITE HEART SUIT	Miscellaneous Symbols
15	👍	0x1f44d	1854	0.812126	213	460	1181	THUMBS UP SIGN	Miscellaneous Symbols
16	😓	0x1f629	1808	0.826234	1069	336	403	WEARY FACE	Emoticons
17	🙄	0x1f64f	1539	0.793848	124	648	767	PERSON WITH FOLDED HA	Emoticons
18	👊	0x270c	1534	0.79048	173	476	685	VICTORY HAND	Dingbats

ภาพที่ 4.6 ข้อมูลการจัดลำดับความรู้สึกจากอิโมจิ

emoji	unicode_code_point	occurrences	position	negative	neutral	positive	unicode_name	unicode_block
🥲	0x1f622	14622	0.805201	3614	4163	6845	FACE WITH TEARS OF JOY	Emoticons
🖤	0x2764	8050	0.746943	355	1334	6361	HEAVY BLACK HEART	Dingbats
🖤	0x2665	7144	0.753866	252	1542	4950	BLACK HEART SUIT	Miscellaneous Symbols
😊	0x1f60d	6359	0.765282	329	1390	4640	SMILING FACE WITH HEART-SHAPED EYES	Emoticons
😭	0x1f62d	5526	0.803352	2412	1218	1896	LOUDLY CRYING FACE	Emoticons
💋	0x1f618	3648	0.85448	193	702	2753	FACE THROWING A KISS	Emoticons
😊	0x1f60a	3186	0.812302	189	754	2243	SMILING FACE WITH SMILING EYES	Emoticons
👌	0x1f44c	2925	0.805223	274	728	1923	OK HAND SIGN	Miscellaneous Symbols and Pictographs
💕	0x1f495	2400	0.765726	99	683	1618	TWO HEARTS	Miscellaneous Symbols and Pictographs
👏	0x1f44f	2326	0.78713	243	634	1459	CLAPPING HANDS SIGN	Miscellaneous Symbols and Pictographs
😬	0x1f601	2189	0.796151	278	648	1263	GRIMACING FACE WITH SMILING EYES	Emoticons
☺	0x263a	2062	0.798834	128	449	1485	WHITE SMILING FACE	Miscellaneous Symbols
💕	0x2661	1975	0.763695	102	448	1425	WHITE HEART SUIT	Miscellaneous Symbols
👍	0x1f44d	1854	0.812126	213	460	1181	THUMBS UP SIGN	Miscellaneous Symbols and Pictographs
😓	0x1f629	1803	0.826214	1069	336	403	WEARY FACE	Emoticons
🙏	0x1f64f	1539	0.793848	124	648	767	PERSON WITH FOLDED HANDS	Emoticons
✌	0x270c	1534	0.79048	173	476	885	VICTORY HAND	Dingbats

ภาพที่ 4.7 ชุดข้อมูลที่ถูกรวบรวมไว้ในฐานข้อมูล MySQL

4. การวิเคราะห์และจำแนกความรู้สึกจากอีโมจิ

หลังจากที่เตรียมข้อความทวิตและชุดข้อมูลที่จะนำมาทำการวิเคราะห์และจำแนกความรู้สึกเรียบร้อยแล้ว ผู้จัดทำจะนำ unicode ที่ได้จากขั้นตอนในข้อ 2. มาทำการแยกอีโมจิแต่ละตัวที่มีอยู่ในข้อความทวิต แล้วแยกการดำเนินการออกเป็นสองกรณี คือ กรณีที่มีเพียงหนึ่งอีโมจิ และกรณีที่มีมากกว่าหนึ่งอีโมจิ หลังจากนั้นก็จะจัดเก็บผลการจำแนกของฐานข้อมูล ตัวอย่างการทำงานของโปรแกรมแสดงซอร์สโค้ดดังภาพที่ 4.8 และผลลัพธ์จากการทำงานของโปรแกรกดังภาพที่ 4.9

```

for row in range(len(unicode)):
    emojis = unicode[row]
    if len(emojis) == 1:
        data_emoji = select_data(emojis)
        if not data_emoji:
            insert_data([str(emojis),0,0,0,0, '-',0])
            continue
        occurrences = int(data_emoji[2])
        pnegative=(int(data_emoji[4])+1)/(occurrences+3)
        pneutral =(int(data_emoji[5])+1)/(occurrences+3)
        ppositive=(int(data_emoji[6])+1)/(occurrences+3)
        sentiment_score = ppositive - pnegative
        neg_neu = -1 + pnegative * 2
        neu_pos = neg_neu + pneutral * 2
    else:
        occurrences = []
        neg = []
        neu = []
        pos = []
        for i in range(len(emojis)):
            data_emoji = select_data(emojis[i])
            if not data_emoji:
                continue
            occurrences.append(int(data_emoji[2]))
            neg.append(int(data_emoji[4]))
            neu.append(int(data_emoji[5]))
            pos.append(int(data_emoji[6]))
        if not occurrences:
            insert_data([str(emojis),0,0,0,0, '-',0])
            continue

```

```

maxx = max(occurrences)
nneg = []
nneu = []
npos = []
for i in range(len(occurrences)):
    nneg.append(neg[i] * maxx / occurrences[i])
    nneu.append(neu[i] * maxx / occurrences[i])
    npos.append(pos[i] * maxx / occurrences[i])
nnegative = sum(nneg) / len(nneg)
nneutral = sum(nneu) / len(nneu)
npositive = sum(npos) / len(npos)
pnegative = (nnegative + 1) / (maxx + 3)
pneutral = (nneutral + 1) / (maxx + 3)
ppositive = (npositive + 1) / (maxx + 3)
sentiment_score = ppositive - pnegative
neg_neu = -1 + pnegative * 2
neu_pos = neg_neu + pneutral * 2
if sentiment_score < neg_neu:
    result = 'negative'
    percentage = pnegative*100
elif sentiment_score <= neu_pos:
    result = 'neutral'

```

ภาพที่ 4.8 ตัวอย่างซอร์สโค้ดในการวิเคราะห์และจำแนกข้อความทวิต

tweetID	unicode_emoji	negative	neutral	positive	score	result	percentage
1	[\ud83c\udf39, \ud83c\udf39, \ud83c\udf39]	0.0174419	0.354651	0.627907	0.610465	positive	62.7907
2	[\ud83d\ude83, \ud83d\ude4c, \ud83d\ude4c]	0.0833333	0.218779	0.697888	0.614555	positive	69.7888
3	[\ud83d\ude22]	0.162162	0.459459	0.378378	0.216216	neutral	45.9459
4	[\ud83d\ude09]	0.0992768	0.337278	0.563445	0.464168	positive	56.3445
5	[\ud83d\ude4c]	0.10093	0.237716	0.661355	0.560425	positive	66.1355
6	[\ud83d\ude4d]	0.114887	0.248112	0.637001	0.522114	positive	63.7001
7	[\ud83c\udfb6, \ud83c\udfb6, \ud83d\udd1e]	0.0982456	0.273684	0.62807	0.529825	positive	62.807
8	[\ud83d\ude99]	0.0317982	0.203947	0.764254	0.732456	positive	76.4254
9	[\ud83e\udd8f, \ud83d\udec11, \ud83e\udd8f]	0.285714	0.619048	0.0952...	-0.190476	neutral	61.9048
10	[\ud83d\ude90, \u2600, \ud83d\ude01, \u270c]	0.10454	0.332669	0.562791	0.45825	positive	56.2791
11	[\ud83c\udfb6, \u2600, \ud83c\udfb6]	0.0708409	0.345064	0.584095	0.513254	positive	58.4095
12	[\ud83d\ude4c, \ud83c\udf79]	0.0915354	0.248688	0.659777	0.568241	positive	65.9777
13	[\u2764, \ud83e\udd17]	0.0440994	0.165714	0.790186	0.746087	positive	79.0186
14	[\ud83d\ude09]	0.0992768	0.337278	0.563445	0.464168	positive	56.3445
15	[\ud83d\ude99, \ud83d\ude97]	0.0545877	0.24971	0.695703	0.641115	positive	69.5703
16	[\ud83e\udd8f, \ud83d\udec11, \ud83e\udd8f]	0.285714	0.619048	0.0952...	-0.190476	neutral	61.9048
17	[\ud83d\udecc]	0.166667	0.5	0.333333	0.166667	neutral	50
18	[\ud83c\udfb6, \ud83c\udfb6]	0.0955777	0.269615	0.634807	0.53923	positive	63.4807
19	[\ud83e\udd14, \ud83e\udd14, \ud83e\udd14, ...]	0.0875	0.565625	0.346875	0.259375	neutral	56.5625
20	[\ud83d\udecfa]	0.0810811	0.432432	0.486486	0.405405	positive	48.6486

ภาพที่ 4.9 ผลลัพธ์จากการวิเคราะห์และจำแนกข้อความทวิต

บทที่ 5

ข้อสรุป

ในบทนี้จะกล่าวถึง ข้อสรุปจากการทดสอบ ปัญหาและอุปสรรค และวิธีการแก้ปัญหา จากการวิเคราะห์ความรู้สึกจากอิโมจิเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์ โดยมีรายละเอียดดังต่อไปนี้

5.1 ข้อสรุป

การวิเคราะห์ความรู้สึกจากอิโมจิเกี่ยวกับการท่องเที่ยวของผู้ใช้ทวิตเตอร์จะจำแนกความรู้สึกของข้อความทวิตออกเป็นความรู้สึกเชิงลบ ความรู้สึกเป็นกลาง หรือความรู้สึกเชิงบวก ซึ่งระบบถูกพัฒนาด้วยภาษาไพธอน ประกอบด้วย 2 กระบวนการหลัก ๆ คือ การเตรียมข้อมูล และการวิเคราะห์ความรู้สึก โดยนำการจัดลำดับความรู้สึกจากอิโมจิ (Emoji Sentiment Ranking) ที่ได้จากงานวิจัยของ P. Kralj Novak และคณะ มาใช้ในการวิเคราะห์และจำแนกความรู้สึกของข้อความทวิตที่เกี่ยวข้องกับการท่องเที่ยว เนื่องจากในปัจจุบันการแสดงความรู้สึกผ่านสื่อ จะไม่เขียนแต่เพียงข้อความที่เป็นตัวอักษรเท่านั้น แต่จะมีการใช้อิโมจิ และรูปภาพประกอบ ดังนั้นการวิเคราะห์และจำแนกความรู้สึกโดยใช้อิโมจิมาประกอบ จะช่วยให้การวิเคราะห์ความรู้สึกมีความใกล้เคียงกับความรู้สึกจริงมากขึ้น ดังนั้นการวิเคราะห์ความรู้สึกของนักท่องเที่ยวจากข้อความทวิตจึงน่าจะเป็นประโยชน์ในด้านการท่องเที่ยว เพราะสามารถนำมาใช้เพื่อประกอบการตัดสินใจเชิงธุรกิจในการพัฒนาและปรับปรุงสถานที่ท่องเที่ยว นั้นต่อไปได้ ผู้จัดทำยังสังเกตเห็นว่า ถ้าหากความน่าจะเป็นที่ข้อความมีความรู้สึกเป็นกลางมีค่าน้อย และความน่าจะเป็นที่ข้อความมีความรู้สึกเป็นลบกับความรู้สึกเป็นบวกมีค่ามากพอ ๆ กัน จะทำให้การวิเคราะห์ความรู้สึกมีค่าความถูกต้องน้อยลง ผู้จัดทำจึงกำหนดให้เป็นข้อความเชิงลบ ซึ่งทำให้ค่าความถูกต้องและประสิทธิภาพโดยรวมเฉลี่ยดีขึ้น

5.2 ปัญหาและอุปสรรค

1. ข้อความทวิตที่นำมาวิเคราะห์เป็นข้อความที่ได้จากการใช้เครื่องมือ Spring XD ในการรวบรวม ไม่ได้มาจากผู้ใช้ทวิตเตอร์ที่เผยแพร่ข้อความทวิตนั้นจริง ๆ ทำให้เราไม่ทราบระดับความรู้สึกที่แท้จริงของผู้ใช้ทวิตเตอร์ว่ามีระดับความรู้สึกเป็นลบ เป็นกลางหรือเป็นบวก
2. ข้อความส่วนใหญ่เป็นข้อความที่มีความรู้สึกเชิงบวกและเป็นกลาง และมีข้อความที่มีความรู้สึกเชิงลบจำนวนน้อยมาก จึงต้องหาข้อความเชิงลบเพิ่มให้เท่ากับข้อความเชิงบวกและข้อความที่เป็นกลาง เพื่อให้ค่าวัดประสิทธิภาพมีความถูกต้องมากที่สุด

5.3 วิธีการแก้ปัญหา

1. ใช้การจำแนกความรู้สึกของข้อความทวิตด้วยผู้จัดทำเอง ซึ่งเป็นวิธีการที่เหมาะสมและน่าจะมีค่าใกล้เคียงกับผลจริงที่สุด
2. เพิ่มขอบเขตของข้อความที่จะนำมาทดสอบประสิทธิภาพ โดยเพิ่มข้อความทวิตเชิงลบตั้งแต่

รายการอ้างอิง

- [1] Emoji Statistics [Online]. 2018. Available from: <https://emojipedia.org/stats/> [2018, November 7]
- [2] Twitter by the Numbers: Stats, Demographics & Fun Facts [Online]. 2018. Available from: <https://www.omnicoreagency.com/twitter-statistics/> [2018, November 6]
- [3] P. Kralj Novak, and others. Emoji Sentiment Ranking [Online]. 2015. Available from: kt.ijs.si/data/Emoji_sentiment_ranking/ [2018, October 4]
- [4] P. Kralj Novak, and others. Sentiment of Emojis (December 2015) : 1-22.
- [5] emoji Meaning in the Cambridge English Dictionary [Online]. Available from: <https://dictionary.cambridge.org/dictionary/english/emoji> [2019, March 24]
- [6] Alex Hern. Don't know the difference between emoji and emoticons? Let me explain [Online]. 2015. Available from: <https://www.theguardian.com/technology/2015/feb/06/difference-between-emoji-and-emoticons-explained> [2019, March 24]
- [7] Nakano, Mamiko. Why and how I created emoji: Interview with Shigetaka Kurita [Online]. 2016. Available from: <https://web.archive.org/web/20160610220635/http://ignition.co/105> [2019, March 24]
- [8] Courtney Seiter. The psychology of emojis [Online]. 2015. Available from: <http://thenextweb.com/insider/2015/06/23/the-psychology-of-emojis/> [2019, March 24]
- [9] Jon Russell. Stickers: From Japanese craze to global mobile messaging phenomenon [Online]. 2013. Available from: <https://thenextweb.com/asia/2013/07/12/stickers/> [2019, March 24]

- [10] Edwina De Abreu. Emoji 101 [Online]. 2015. Available from: <https://www.ovrdrv.com/emoji-101/> [2019, March 24]
- [11] Python Tutorial [Online]. Available from: <https://www.w3schools.com/python/default.asp> [2019, March 24]
- [12] Wikipedia. Python (programming language) [Online]. Available from: [https://en.wikipedia.org/wiki/Python_\(programming_language\)](https://en.wikipedia.org/wiki/Python_(programming_language)) [2019, March 24]
- [13] Oracle Corporation. What is MySQL? [Online]. Available from: <https://dev.mysql.com/doc/refman/8.0/en/what-is-mysql.html> [2019, March 24]
- [14] Amornthep Thongchiw. Not only using Accuracy on Model Evaluation [Online]. 2018. Available from: <https://thaidatascience.wordpress.com/2018/04/11/not-only-using-accuracy-on-model-evaluation/> [2019, April 30]
- [15] Mark Fisher, and others. Spring XD Guide [Online]. 2018. Available from: <https://docs.spring.io/spring-xd/docs/current-SNAPSHOT/reference/html/> [2019, May 6]

ภาคผนวก

ภาคผนวก ก

แบบเสนอหัวข้อโครงการ รายวิชา 2301399 Project Proposal

ปีการศึกษา 2561

ชื่อโครงการ (ภาษาไทย)	การวิเคราะห์ความรู้สึกของผู้ใช้ทวีตเตอร์จากอีโมจิ
ชื่อโครงการ (ภาษาอังกฤษ)	Twitter User's Sentiment Analysis from Emoji
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.รัชลิดา ลิปิกรณ์
ผู้ดำเนินการ	นางสาวกนกกานต์ ด่านสว่างกุล เลขประจำตัวนิสิต 5833603023 สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์และวิทยาการ คอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

หลักการและเหตุผล

ในปัจจุบันผู้คนมีการใช้งานสื่อสังคม (social media) เพิ่มขึ้นเรื่อยๆ เพื่อติดต่อสื่อสารและแลกเปลี่ยนข้อมูลต่าง ๆ เช่น ความคิดเห็น ความรู้สึก หรือประสบการณ์ โดยจะแสดงในรูปแบบของข้อความ รูปภาพ วิดีโอ หรืออีโมจิ เป็นต้น และมีงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความรู้สึกจากข้อความเป็นจำนวนมาก แต่การสื่อสารบนสื่อสังคมด้วยข้อความเพียงอย่างเดียวอาจจะไม่เพียงพอที่จะวิเคราะห์ได้อย่างถูกต้องว่าผู้ใช้งานที่แสดงข้อความนั้นมีความรู้สึกอย่างไรกับข้อความนั้น หรือผู้ใช้งานที่อ่านข้อความนั้นอาจเข้าใจข้อความนั้นด้วยความรู้สึกที่ตรงกันข้ามกับผู้แสดงข้อความนั้นออกมา อีโมจิจะช่วยให้เข้าใจความรู้สึกของผู้ที่แสดงข้อความนั้นได้แม่นยำมากขึ้น เนื่องจากอีโมจิเป็นอักขรภาพที่สามารถสื่อสารได้ดีจากสีหน้าของอีโมจิหรือสัญลักษณ์ที่คนส่วนใหญ่เข้าใจความหมายตรงกัน และในปัจจุบันมีอีโมจิมากถึง 2,823 อีโมจิที่อยู่ในมาตรฐานยูนิโคด [2] ที่ผู้ใช้สามารถเลือกใช้ได้หลากหลายและสามารถสื่อความรู้สึกได้ตรงความต้องการมากขึ้น อีโมจิจึงเป็นข้อมูลรูปแบบหนึ่งที่น่าสนใจที่จะนำมาศึกษา

ทวีตเตอร์เป็นสื่อสังคมที่นิยมใช้กันอย่างกว้างขวาง Omnicore Agency [1] ได้เปิดเผยจำนวนผู้ใช้ทวีตเตอร์ว่ามีประมาณ 100 ล้านบัญชีที่มีการใช้งานทุกวัน ทวีตเตอร์ยังเป็นแหล่งข้อมูลที่มีจำนวนมากมหาศาลและมีเนื้อหาหลากหลายทั้งข้อมูลทางด้านเทคโนโลยี การเมืองการปกครอง เศรษฐกิจ ความบันเทิง รวมถึงข้อมูลเกี่ยวกับการท่องเที่ยว ซึ่งเป็นข้อมูลที่สามารถส่งผลกระทบต่ออุตสาหกรรมการท่องเที่ยวในแต่ละประเทศได้ ทั้งนี้ ถ้าหากมีผู้ใช้ทวีตเตอร์แสดงความคิดเห็นต่อสถานที่ท่องเที่ยวในเชิงบวก จะช่วยส่งเสริมการท่องเที่ยวให้กับสถานที่ท่องเที่ยวนั้น และผู้ใช้ที่เห็นความคิดเห็นนี้อาจจะมีการแบ่งปันข้อมูลส่งต่อไปให้กับผู้ใช้คนอื่น ๆ ให้รู้จักกันไปในวงกว้าง ในทางกลับกัน ถ้าหากมีผู้ใช้ทวีตเตอร์แสดงความคิดเห็นต่อสถานที่

ท่องเที่ยวในเชิงลบ ผู้ที่เป็นเจ้าของสถานที่หรือผู้ที่มีหน้าที่ในการรับผิดชอบต่อการดูแลสถานที่นั้น ๆ สามารถรับรู้ผลสะท้อนกลับ (feedback) จากการให้บริการของสถานที่ท่องเที่ยวที่ตน และสืบหาวิธีการปรับปรุงและพัฒนาการให้บริการของสถานที่ท่องเที่ยวที่ตน ๆ ได้จากความคิดเห็นเพิ่มเติมได้ จะช่วยให้เกิดประโยชน์ต่อสถานที่ท่องเที่ยวที่ตน และยังเป็นการสร้างรายได้ให้กับประเทศ หรือบริเวณสถานที่ท่องเที่ยวที่ตนด้วย

ผู้จัดทำจึงสนใจที่จะพัฒนาระบบวิเคราะห์ความรู้สึกของผู้ใช้ทวิตเตอร์จากอิโมจิที่เป็นรูปแบบข้อมูลที่ยังไม่ค่อยมีผู้นำมาศึกษาและวิเคราะห์ ซึ่งจะนำข้อมูลเกี่ยวกับการท่องเที่ยวในทวิตเตอร์มาใช้เป็นกรณีศึกษา โดยจะนำ Emoji Sentiment Ranking [3] ที่ได้จากงานวิจัยของ P. Kralj Novak และคณะ [4] ซึ่งทำการวิเคราะห์ความรู้สึกและเสนอออกมาในรูปแบบของแถบความรู้สึก (Sentiment bar) มาใช้เป็นแหล่งอ้างอิง และแบ่งความรู้สึกออกเป็น 3 แบบ ได้แก่ ความรู้สึกเชิงบวก (Positive) ความรู้สึกเชิงลบ (Negative) และความรู้สึกเป็นกลาง (Neutral)

วัตถุประสงค์

เพื่อพัฒนาระบบวิเคราะห์ความรู้สึกของผู้ใช้ทวิตเตอร์จากอิโมจิที่ผู้ใช้งานนั้นแสดงความคิดเห็นว่ามีความคิดเห็นเป็นบวก เป็นลบ หรือเป็นกลาง โดยใช้ข้อมูลเกี่ยวกับการท่องเที่ยวปีการศึกษา

ขอบเขตของโครงการ

1. วิเคราะห์ความรู้สึกจากอิโมจิในแต่ละทวิตเท่านั้น
2. วิเคราะห์อิโมจิแค่ 751 อิโมจิเท่านั้น
3. ความรู้สึกของผู้ใช้งานจะถูกแบ่งเป็น 3 แบบ คือ เป็นบวก เป็นลบ หรือเป็นกลาง
4. ใช้ข้อมูลจากทวิตเตอร์เกี่ยวกับการท่องเที่ยวที่ประกอบด้วย #travel และอยู่ในปี พ.ศ. 2560 เป็นกรณีศึกษาสำหรับโครงการนี้เท่านั้น
5. ภาษาที่ใช้ในการสื่อสารผ่านทวิตเตอร์เป็นภาษาอังกฤษเท่านั้น
6. รหัสอิโมจิเป็นรหัสมาตรฐานบนระบบปฏิบัติการ Windows, IOS และ Android

วิธีการดำเนินงาน

ก.แผนการศึกษา

1. ศึกษาค้นคว้าข้อมูลที่เกี่ยวข้องกับการวิเคราะห์อิโมจิและความรู้สึก
2. ศึกษาเครื่องมือ โปรแกรมและเทคนิคที่ใช้ในโครงการ
3. กำหนดขอบเขตและวิธีการดำเนินงานของโครงการ
4. วิเคราะห์และออกแบบระบบวิเคราะห์ความรู้สึกของผู้ใช้ทวิตเตอร์จากอิโมจิ

5. พัฒนาระบบวิเคราะห์ความรู้สึกของผู้ใช้ทวีตเตอร์จากอิโมจิ
6. ทดสอบระบบวิเคราะห์ความรู้สึกของผู้ใช้ทวีตเตอร์จากอิโมจิ
7. ปรับปรุงและแก้ไขระบบวิเคราะห์ความรู้สึกของผู้ใช้ทวีตเตอร์จากอิโมจิ
8. สรุปผลการดำเนินการ
9. จัดทำเอกสาร

ข. ระยะเวลาที่ศึกษา

ขั้นตอนการดำเนินงาน	พ.ศ. 2561					พ.ศ. 2562			
	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.
1. ศึกษาค้นคว้าข้อมูลที่เกี่ยวข้อง									
2. ศึกษาเครื่องมือ โปรแกรมและเทคนิคที่ใช้									
3. กำหนดขอบเขตและวิธีการดำเนินงานของโครงการ									
4. วิเคราะห์และออกแบบระบบ									
5. พัฒนาระบบ									
6. ทดสอบระบบ									
7. ปรับปรุงและแก้ไขระบบ									
8. สรุปผลการดำเนินการ									
9. จัดทำเอกสาร									

ประโยชน์ที่คาดว่าจะได้รับ

ก.ด้านความรู้และประสบการณ์ต่อตัวนิสิต

1. ได้ความรู้และความเข้าใจเกี่ยวกับวิธีการวิเคราะห์ความรู้สึกจากอิโมจิ
2. ได้เรียนรู้การใช้โปรแกรมที่นำมาใช้ในการพัฒนาซอฟต์แวร์
3. ได้เรียนรู้การวางแผนการทำงาน

ข. ประโยชน์ที่ได้จากโครงการที่พัฒนาขึ้น

1. ทำให้ทราบถึงความรู้สึกของผู้ใช้งานที่แสดงความคิดเห็นด้วยอิโมจิ
2. สามารถนำไปใช้ประกอบการตัดสินใจในการพัฒนาสถานที่ท่องเที่ยว
3. สามารถนำไปใช้ในการพัฒนาระบบอื่น ๆ ที่เกี่ยวกับการวิเคราะห์ความรู้สึกต่อไปได้

อุปกรณ์และเครื่องมือที่ใช้

ก. ฮาร์ดแวร์

1. คอมพิวเตอร์ที่มี CPU ความเร็วอย่างน้อย 2.0 GHz และ RAM อย่างน้อย 8 GB

ข. ซอฟต์แวร์

1. ระบบปฏิบัติการ Windows
2. Python เวอร์ชัน 3.7 หรือสูงกว่า

งบประมาณ

External drive 5000.00 บาท

เอกสารอ้างอิง

- [1] Twitter by the Numbers: Stats, Demographics & Fun Facts [Online]. 2018. Available from: <https://www.omnicoreagency.com/twitter-statistics/> [2018, November 6]
- [2] Emoji Statistics [Online]. 2018. Available from: <https://emojipedia.org/stats/> [2018, November 7]
- [3] P. Kralj Novak, and others. Emoji Sentiment Ranking [Online]. 2015. Available from: kt.ijs.si/data/Emoji_sentiment_ranking/ [2018, October 4]
- [4] P. Kralj Novak, and others. Sentiment of Emojis (December 2015) : 1-22.

ภาคผนวก ข










อิโมจิที่นำมาใช้ในการวิเคราะห์




emoji	unicode_code_point	occurrences	position	negative	neutral	positive	unicode_name	unicode_block
😭	0x1f602	14622	0.805101	3614	4163	6845	FACE WITH TEARS OF JOY	Emoticons
❤️	0x2764	8050	0.746943	355	1334	6361	HEAVY BLACK HEART	Dingbats
♥️	0x2665	7144	0.753806	252	1942	4950	BLACK HEART SUIT	Miscellaneous Sy...
😊	0x1f60d	6359	0.765292	329	1390	4640	SMILING FACE WITH HEART...	Emoticons
😭	0x1f62d	5526	0.803352	2412	1218	1896	LOUDLY CRYING FACE	Emoticons
😘	0x1f618	3648	0.85448	193	702	2753	FACE THROWING A KISS	Emoticons
😊	0x1f60a	3186	0.813302	189	754	2243	SMILING FACE WITH SMILIN...	Emoticons
👌	0x1f44c	2925	0.805223	274	728	1923	OK HAND SIGN	Miscellaneous Sy...
💕	0x1f495	2400	0.765726	99	683	1618	TWO HEARTS	Miscellaneous Sy...
👐	0x1f44f	2336	0.78713	243	634	1459	CLAPPING HANDS SIGN	Miscellaneous Sy...
😄	0x1f601	2189	0.796151	278	648	1263	GRINNING FACE WITH SMILI...	Emoticons
😊	0x263a	2062	0.798634	128	449	1485	WHITE SMILING FACE	Miscellaneous Sy...
☹️	0x2661	1975	0.763695	102	448	1425	WHITE HEART SUIT	Miscellaneous Sy...
👍	0x1f44d	1854	0.812126	213	460	1181	THUMBS UP SIGN	Miscellaneous Sy...
😓	0x1f629	1808	0.826214	1069	336	403	WEARY FACE	Emoticons
🙌	0x1f64f	1539	0.793848	124	648	767	PERSON WITH FOLDED HANDS	Emoticons
✌️	0x270c	1534	0.79048	173	476	885	VICTORY HAND	Dingbats
😏	0x1f60f	1522	0.764977	170	676	676	SMIRKING FACE	Emoticons
😊	0x1f609	1521	0.844833	151	513	857	WINKING FACE	Emoticons
🙌	0x1f64c	1506	0.7906	152	358	996	PERSON RAISING BOTH HAN...	Emoticons
🙈	0x1f648	1456	0.738881	238	350	868	SEE-NO-EVIL MONKEY	Emoticons
💪	0x1f4aa	1409	0.806704	101	424	884	FLEXED BICEPS	Miscellaneous Sy...
😊	0x1f604	1398	0.794973	191	426	781	SMILING FACE WITH OPEN ...	Emoticons
😐	0x1f612	1385	0.857621	819	266	300	UNAMUSED FACE	Emoticons
💃	0x1f483	1344	0.852765	59	237	1048	DANCER	Miscellaneous Sy...
💖	0x1f496	1263	0.762239	54	254	955	SPARKLING HEART	Miscellaneous Sy...
😊	0x1f603	1206	0.734782	86	361	759	SMILING FACE WITH OPEN ...	Emoticons
😞	0x1f614	1205	0.866146	559	263	383	PENSIVE FACE	Emoticons
😱	0x1f631	1130	0.773313	298	319	513	FACE SCREAMING IN FEAR	Emoticons
🎉	0x1f389	1125	0.743636	43	207	875	PARTY POPPER	Miscellaneous Sy...
😓	0x1f61c	1035	0.822247	115	333	587	FACE WITH STUCK-OUT TO...	Emoticons
☯️	0x262f	992	0.383524	5	981	6	YIN YANG	Miscellaneous Sy...
🌸	0x1f338	946	0.570022	37	255	654	CHERRY BLOSSOM	Miscellaneous Sy...
💜	0x1f49c	939	0.73087	41	241	657	PURPLE HEART	Miscellaneous Sy...
💙	0x1f499	912	0.703048	29	186	697	BLUE HEART	Miscellaneous Sy...
✨	0x2728	848	0.56043	43	463	342	SPARKLES	Dingbats
😬	0x1f633	846	0.797186	277	277	292	FLUSHED FACE	Emoticons
💚	0x1f497	836	0.799726	42	201	593	GROWING HEART	Miscellaneous Sy...
★	0x2605	828	0.353317	25	543	260	BLACK STAR	Miscellaneous Sy...
■	0x2588	798	0.634172	71	682	45	FULL BLOCK	Block Elements
☀️	0x2600	786	0.545531	21	377	388	BLACK SUN WITH RAYS	Miscellaneous Sy...
😜	0x1f621	756	0.861522	403	81	272	POUTING FACE	Emoticons
😊	0x1f60e	754	0.765644	79	224	451	SMILING FACE WITH SUNGL...	Emoticons
😭	0x1f622	749	0.813897	288	168	293	CRYING FACE	Emoticons
💋	0x1f48b	734	0.800581	28	169	537	KISS MARK	Miscellaneous Sy...
😊	0x1f60b	734	0.764985	33	203	498	FACE SAVOURING DELICIOU...	Emoticons
🙈	0x1f64a	725	0.740523	97	197	431	SPEAK-NO-EVIL MONKEY	Emoticons
😴	0x1f634	718	0.849923	303	170	245	SLEEPING FACE	Emoticons
🎵	0x1f3b6	701	0.797256	67	189	445	MULTIPLE MUSICAL NOTES	Miscellaneous Sy...
💞	0x1f49e	687	0.828706	27	123	537	REVOLVING HEARTS	Miscellaneous Sy...
😊	0x1f60c	665	0.847121	93	157	415	RELIEVED FACE	Emoticons
💧	0x1f525	651	0.61555	80	400	171	FIRE	Miscellaneous Sy...
💯	0x1f4af	637	0.871819	179	202	256	HUNDRED POINTS SYMBOL	Miscellaneous Sy...
🔫	0x1f52b	604	0.893761	298	126	180	PISTOL	Miscellaneous Sy...

❤	0x1f49b	602	0.751632	25	123	454	YELLOW HEART	Miscellaneous Sy...
👤	0x1f481	549	0.839629	105	159	285	INFORMATION DESK PERSON	Miscellaneous Sy...
♥	0x1f49a	537	0.737499	41	101	395	GREEN HEART	Miscellaneous Sy...
🎵	0x266b	533	0.396607	33	313	187	BEAMED EIGHTH NOTES	Miscellaneous Sy...
😞	0x1f61e	532	0.825188	255	85	192	DISAPPOINTED FACE	Emoticons
😊	0x1f606	527	0.809376	81	148	298	SMILING FACE WITH OPEN ...	Emoticons
😬	0x1f61d	496	0.808809	65	155	276	FACE WITH STUCK-OUT TO...	Emoticons
😪	0x1f62a	482	0.85822	208	105	169	SLEEPY FACE	Emoticons
🔠	0xffffd	472	0.112933	78	275	119	REPLACEMENT CHARACTER	Specials
😓	0x1f62b	467	0.839518	227	81	159	TIRED FACE	Emoticons
😊	0x1f605	462	0.827655	135	109	218	SMILING FACE WITH OPEN ...	Emoticons
👊	0x1f44a	458	0.824516	121	111	226	FISTED HAND SIGN	Miscellaneous Sy...
👤	0x1f480	456	0.739462	214	123	119	SKULL	Miscellaneous Sy...
😄	0x1f600	439	0.786285	37	114	288	GRINNING FACE	Emoticons
😘	0x1f61a	424	0.853415	20	81	323	KISSING FACE WITH CLOSE...	Emoticons
😺	0x1f63b	417	0.729666	28	101	288	SMILING CAT FACE WITH HE...	Emoticons
©	0xa9	416	0.740148	54	259	103	COPYRIGHT SIGN	Latin-1 Supplement
👁	0x1f440	410	0.759422	93	198	119	EYES	Miscellaneous Sy...
💖	0x1f498	395	0.785037	18	87	290	HEART WITH ARROW	Miscellaneous Sy...
🐔	0x1f413	384	0.490808	41	291	52	ROOSTER	Miscellaneous Sy...
👋	0x1f44b	382	0.797944	60	103	219	WAVING HAND SIGN	Miscellaneous Sy...
☕	0x2615	382	0.685215	53	182	147	HOT BEVERAGE	Miscellaneous Sy...
✋	0x270b	378	0.838292	124	82	172	RAISED HAND	Dingbats
🎉	0x1f38a	363	0.723622	20	59	284	CONFETTI BALL	Miscellaneous Sy...
🍕	0x1f355	352	0.544581	19	166	167	SLICE OF PIZZA	Miscellaneous Sy...
❄	0x2744	349	0.760182	34	103	212	SNOWFLAKE	Dingbats
😞	0x1f625	341	0.881977	108	83	150	DISAPPOINTED BUT RELIEVE...	Emoticons
😵	0x1f615	340	0.856712	205	66	69	CONFUSED FACE	Emoticons
💥	0x1f4a5	329	0.587206	23	234	72	COLLISION SYMBOL	Miscellaneous Sy...
💔	0x1f494	328	0.835832	136	96	96	BROKEN HEART	Miscellaneous Sy...
🏆	0x1f624	327	0.884008	157	82	88	FACE WITH LOOK OF TRIUMPH	Emoticons
😠	0x1f608	325	0.762498	66	106	153	SMILING FACE WITH HORNS	Emoticons
▶	0x25ba	325	0.516489	40	192	93	BLACK RIGHT-POINTING POI...	Geometric Shapes
✈	0x2708	322	0.52194	13	161	148	AIRPLANE	Dingbats
⬆	0x1f51d	303	0.798434	26	106	171	TOP WITH UPWARDS ARRO...	Miscellaneous Sy...
😬	0x1f630	302	0.919654	132	44	126	FACE WITH OPEN MOUTH A...	Emoticons
😐	0x1f611	299	0.844273	154	85	60	EXPRESSIONLESS FACE	Emoticons
⚽	0x26bd	299	0.193496	15	83	201	SOCCER BALL	Miscellaneous Sy...
👑	0x1f451	298	0.698805	12	65	221	CROWN	Miscellaneous Sy...
😿	0x1f639	295	0.807414	88	77	130	CAT FACE WITH TEARS OF J...	Emoticons
👉	0x1f449	292	0.559176	20	137	135	WHITE RIGHT POINTING BA...	Miscellaneous Sy...
🍃	0x1f343	291	0.572324	29	122	140	LEAF FLUTTERING IN WIND	Miscellaneous Sy...
📦	0x1f381	288	0.709181	11	45	232	WRAPPED PRESENT	Miscellaneous Sy...
😡	0x1f620	288	0.849218	163	49	76	ANGRY FACE	Emoticons
🐧	0x1f427	284	0.540409	11	131	142	PENGUIN	Miscellaneous Sy...
☆	0x2606	282	0.563398	8	144	130	WHITE STAR	Miscellaneous Sy...
🍀	0x1f340	278	0.768496	6	186	86	FOUR LEAF CLOVER	Miscellaneous Sy...
🎈	0x1f388	277	0.718504	4	68	205	BALLOON	Miscellaneous Sy...
👴	0x1f385	274	0.800787	7	172	95	FATHER CHRISTMAS	Miscellaneous Sy...
😓	0x1f613	273	0.844887	118	59	96	FACE WITH COLD SWEAT	Emoticons
😊	0x1f623	271	0.903308	147	35	89	PERSEVERING FACE	Emoticons
😐	0x1f610	270	0.882733	150	76	44	NEUTRAL FACE	Emoticons
✊	0x270a	270	0.858609	37	79	154	RAISED FIST	Dingbats
😨	0x1f628	269	0.813688	117	73	79	FEARFUL FACE	Emoticons
😵	0x1f616	268	0.868927	130	50	88	CONFOUNDED FACE	Emoticons
zzz	0x1f4a4	267	0.800462	38	91	138	SLEEPING SYMBOL	Miscellaneous Sy...
💓	0x1f493	259	0.783045	15	55	189	BEATING HEART	Miscellaneous Sy...
👎	0x1f44e	258	0.914491	128	51	79	THUMBS DOWN SIGN	Miscellaneous Sy...
💦	0x1f4a6	252	0.779044	35	62	155	SPLASHING SWEAT SYMBOL	Miscellaneous Sy...
✓	0x2714	249	0.565237	31	119	99	HEAVY CHECK MARK	Dingbats
😷	0x1f637	246	0.840377	117	54	75	FACE WITH MEDICAL MASK	Emoticons
⚡	0x26a1	246	0.778781	10	182	54	HIGH VOLTAGE SIGN	Miscellaneous Sy...
🎄	0x1f384	236	0.632526	15	79	142	CHRISTMAS TREE	Miscellaneous Sy...
😊	0x1f64b	236	0.7325	30	60	146	HAPPY PERSON RAISING ON...	Emoticons
💩	0x1f4a9	229	0.83972	94	68	67	PILE OF POO	Miscellaneous Sy...
🎵	0x1f3b5	223	0.71527	23	64	136	MUSICAL NOTE	Miscellaneous Sy...

➡	0x27a1	222	0.47833	2	185	35	BLACK RIGHTWARDS ARROW	Dingbats
😬	0x1f61b	220	0.734419	18	50	152	FACE WITH STUCK-OUT TO...	Emoticons
😬	0x1f62c	214	0.79368	57	58	99	GRIMACING FACE	Emoticons
👩	0x1f46f	211	0.742044	24	69	118	WOMAN WITH BUNNY EARS	Miscellaneous Sy...
💎	0x1f48e	209	0.694055	11	68	130	GEM STONE	Miscellaneous Sy...
🌿	0x1f33f	208	0.600121	1	125	82	HERB	Miscellaneous Sy...
🎂	0x1f382	201	0.748	16	44	141	BIRTHDAY CAKE	Miscellaneous Sy...
★	0x1f31f	199	0.654511	11	111	77	GLOWING STAR	Miscellaneous Sy...
🔮	0x1f52e	199	0.333162	6	133	60	CRYSTAL BALL	Miscellaneous Sy...
!	0x2757	198	0.620121	31	116	51	HEAVY EXCLAMATION MARK...	Dingbats
👨	0x1f46b	197	0.664271	43	60	94	MAN AND WOMAN HOLDING...	Miscellaneous Sy...
🏆	0x1f3c6	194	0.705937	6	39	149	TROPHY	Miscellaneous Sy...
×	0x2716	193	0.514517	8	116	69	HEAVY MULTIPLICATION X	Dingbats
😘	0x1f619	191	0.875675	3	34	154	KISSING FACE WITH SMILIN...	Emoticons
☝	0x261d	191	0.759773	27	77	87	WHITE UP POINTING INDEX	Miscellaneous Sy...
👶	0x26c4	191	0.738725	14	62	115	SNOWMAN WITHOUT SNOW	Miscellaneous Sy...
👅	0x1f445	190	0.738179	23	55	112	TONGUE	Miscellaneous Sy...
🎵	0x266a	190	0.696314	15	57	118	EIGHTH NOTE	Miscellaneous Sy...
🍂	0x1f342	189	0.634541	6	72	111	FALLEN LEAF	Miscellaneous Sy...
💋	0x1f48f	185	0.728952	33	46	106	KISS	Miscellaneous Sy...
🍡	0x1f52a	183	0.764046	38	94	51	HOCHO	Miscellaneous Sy...
🌴	0x1f334	178	0.67589	13	57	108	PALM TREE	Miscellaneous Sy...
👉	0x1f448	174	0.701409	14	71	89	WHITE LEFT POINTING BAC...	Miscellaneous Sy...
🌹	0x1f339	172	0.664431	3	61	108	ROSE	Miscellaneous Sy...
👌	0x1f646	171	0.839917	15	54	102	FACE WITH OK GESTURE	Emoticons
➡	0x279c	170	0.558507	8	126	36	HEAVY ROUND-TIPPED RIGH...	Dingbats
👻	0x1f47b	168	0.740824	28	73	67	GHOST	Miscellaneous Sy...
💰	0x1f4b0	168	0.827466	26	73	69	MONEY BAG	Miscellaneous Sy...
🍻	0x1f37b	165	0.745307	17	45	103	CLINKING BEER MUGS	Miscellaneous Sy...
😬	0x1f645	165	0.838872	76	47	42	FACE WITH NO GOOD GEST...	Emoticons
☀	0x1f31e	162	0.709366	3	64	95	SUN WITH FACE	Miscellaneous Sy...
🍁	0x1f341	161	0.623982	5	72	84	MAPLE LEAF	Miscellaneous Sy...
▪	0x25aa	159	0.145167	15	97	47	BLACK SMALL SQUARE	Geometric Shapes
★	0x2b50	159	0.600764	5	55	99	WHITE MEDIUM STAR	Miscellaneous Sy...
🎀	0x1f380	156	0.651799	6	44	106	RIBBON	Miscellaneous Sy...
▭	0x2501	156	0.54025	14	100	42	BOX DRAWINGS HEAVY HOR...	Box Drawing
☷	0x2637	154	0.0124078	2	140	12	TRIGRAM FOR EARTH	Miscellaneous Sy...
🐷	0x1f437	152	0.602682	11	73	68	PIG FACE	Miscellaneous Sy...
🌺	0x1f33a	150	0.629473	2	62	86	HIBISCUS	Miscellaneous Sy...
🐵	0x1f649	150	0.757695	26	47	77	HEAR-NO-EVIL MONKEY	Emoticons
💅	0x1f485	149	0.755748	27	36	86	NAIL POLISH	Miscellaneous Sy...
🌘	0x1f31a	148	0.759895	23	32	93	NEW MOON WITH FACE	Miscellaneous Sy...
🐶	0x1f436	148	0.654156	12	37	99	DOG FACE	Miscellaneous Sy...
👽	0x1f47d	146	0.612457	13	73	60	EXTRATERRESTRIAL ALIEN	Miscellaneous Sy...
🎙	0x1f3a4	144	0.670892	17	40	87	MICROPHONE	Miscellaneous Sy...
👩	0x1f46d	144	0.667224	20	36	88	TWO WOMEN HOLDING HANDS	Miscellaneous Sy...
🎧	0x1f3a7	142	0.842971	18	46	78	HEADPHONE	Miscellaneous Sy...
🍸	0x1f378	138	0.662604	12	38	88	COCKTAIL GLASS	Miscellaneous Sy...
👉	0x1f446	138	0.787464	16	60	62	WHITE UP POINTING BACKH...	Miscellaneous Sy...
🍷	0x1f377	137	0.603672	7	68	62	WINE GLASS	Miscellaneous Sy...
®	0xae	137	0.353085	9	80	48	REGISTERED SIGN	Latin-1 Supplement
🍉	0x1f349	136	0.647779	10	33	93	WATERMELON	Miscellaneous Sy...
🏃	0x1f3c3	135	0.735256	12	55	68	RUNNER	Miscellaneous Sy...
😇	0x1f607	135	0.781831	9	36	90	SMILING FACE WITH HALO	Emoticons
☑	0x2611	135	0.690062	2	117	16	BALLOT BOX WITH CHECK	Miscellaneous Sy...
😭	0x1f63f	134	0.811561	78	29	27	CRYING CAT FACE	Emoticons
▭	0x2502	134	0.48318	0	87	47	BOX DRAWINGS LIGHT VERT...	Box Drawing
🍺	0x1f37a	131	0.681566	8	49	74	BEER MUG	Miscellaneous Sy...
💣	0x1f4a3	131	0.846727	40	50	41	BOMB	Miscellaneous Sy...
▶	0x25b6	131	0.498359	7	89	35	BLACK RIGHT-POINTING TRI...	Geometric Shapes
😲	0x1f632	129	0.85362	50	38	41	ASTONISHED FACE	Emoticons
🎸	0x1f3b8	125	0.605443	1	57	67	GIUITAR	Miscellaneous Sy...
🍹	0x1f379	123	0.764165	5	30	88	TROPICAL DRINK	Miscellaneous Sy...
🌀	0x1f4ab	121	0.614438	4	51	66	DIZZY SYMBOL	Miscellaneous Sy...
📖	0x1f4da	119	0.691632	22	34	63	BOOKS	Miscellaneous Sy...

☹️	0x1f636	117	0.868668	50	34	33	FACE WITHOUT MOUTH	Emoticons
🌷	0x1f337	116	0.623398	0	52	64	TULIP	Miscellaneous Sy...
💞	0x1f49d	115	0.77441	8	23	84	HEART WITH RIBBON	Miscellaneous Sy...
➡️	0x1f4a8	115	0.711397	12	46	57	DASH SYMBOL	Miscellaneous Sy...
🏈	0x1f3c8	114	0.823011	7	38	69	AMERICAN FOOTBALL	Miscellaneous Sy...
💍	0x1f48d	112	0.807521	16	25	71	RING	Miscellaneous Sy...
👸	0x1f478	111	0.654472	6	30	75	PRINCESS	Miscellaneous Sy...
☔	0x2614	111	0.687353	21	36	54	UMBRELLA WITH RAIN DROPS	Miscellaneous Sy...
Ⓔ	0x1f1ea	109	0.698877	2	36	71	REGIONAL INDICATOR SYM...	Enclosed Alphanu...
⬜	0x2591	108	0.439438	25	63	20	LIGHT SHADE	Block Elements
🥞	0x1f369	107	0.708816	15	35	57	DOUGHNUT	Miscellaneous Sy...
👽	0x1f47e	105	0.551881	1	64	40	ALIEN MONSTER	Miscellaneous Sy...
🌻	0x1f33b	104	0.607167	1	41	62	SUNFLOWER	Miscellaneous Sy...
☁️	0x2601	104	0.595923	14	43	47	CLOUD	Miscellaneous Sy...
😵	0x1f635	103	0.808814	33	28	42	DIZZY FACE	Emoticons
📖	0x1f4d2	102	0.473945	0	98	4	LEDGER	Miscellaneous Sy...
↑	0x21bf	102	0.710504	0	34	68	UPWARDS HARPOON WITH ...	Arrows
🐅	0x1f42f	100	0.549562	4	43	53	TIGER FACE	Miscellaneous Sy...
👶	0x1f47c	99	0.683211	17	31	51	BABY ANGEL	Miscellaneous Sy...
🍔	0x1f354	98	0.636878	16	38	44	HAMBURGER	Miscellaneous Sy...
😸	0x1f638	97	0.782036	11	34	52	GRINNING CAT FACE WITH ...	Emoticons
👶	0x1f476	96	0.745049	13	27	56	BABY	Miscellaneous Sy...
↑	0x21be	96	0.707143	0	32	64	UPWARDS HARPOON WITH ...	Arrows
🌊	0x1f30a	95	0.779832	8	30	57	WATER WAVE	Miscellaneous Sy...
🍦	0x1f366	95	0.630276	13	24	58	SOFT ICE CREAM	Miscellaneous Sy...
🌸	0x1f490	95	0.727587	3	17	75	BOUQUET	Miscellaneous Sy...
🍓	0x1f353	94	0.689469	3	23	68	STRAWBERRY	Miscellaneous Sy...
👇	0x1f447	94	0.557021	13	44	37	WHITE DOWN POINTING BA...	Miscellaneous Sy...
🍴	0x1f374	92	0.580198	4	33	55	FORK AND KNIFE	Miscellaneous Sy...
👩	0x1f486	92	0.814024	24	23	45	FACE MASSAGE	Miscellaneous Sy...
😓	0x1f627	92	0.77499	36	26	30	ANGUISHED FACE	Emoticons
Ⓔ	0x1f1f8	91	0.654692	8	27	56	REGIONAL INDICATOR SYM...	Enclosed Alphanu...
😊	0x1f62e	90	0.774237	16	33	41	FACE WITH OPEN MOUTH	Emoticons
⬛	0x2593	90	0.658247	2	85	3	DARK SHADE	Block Elements
🌈	0x1f308	88	0.671884	5	31	52	RAINBOW	Miscellaneous Sy...
😸	0x1f63d	88	0.754243	11	14	63	KISSING CAT FACE WITH CL...	Emoticons
😸	0x1f640	88	0.790729	14	30	44	WEARY CAT FACE	Emoticons
🚫	0x1f6ab	88	0.702938	48	32	8	NO ENTRY SIGN	Transport and Ma...
⚠️	0x26a0	88	0.489085	16	62	10	WARNING SIGN	Miscellaneous Sy...
🎮	0x1f3ae	86	0.46743	8	32	46	VIDEO GAME	Miscellaneous Sy...
🍆	0x1f346	84	0.718503	6	37	41	AUBERGINE	Miscellaneous Sy...
🍰	0x1f370	84	0.782506	13	19	52	SHORTCAKE	Miscellaneous Sy...
📦	0x256f	84	0.506911	19	47	18	BOX DRAWINGS LIGHT ARC ...	Box Drawing
✓	0x2713	84	0.627536	9	41	34	CHECK MARK	Dingbats
🍟	0x1f35f	83	0.650434	14	29	40	FRENCH FRIES	Miscellaneous Sy...
👐	0x1f450	83	0.875981	33	19	31	OPEN HANDS SIGN	Miscellaneous Sy...
🙇	0x1f647	83	0.794294	24	23	36	PERSON BOWING DEEPLY	Emoticons
🍌	0x1f34c	82	0.540634	10	25	47	BANANA	Miscellaneous Sy...
🤝	0x1f46c	82	0.62224	29	29	24	TWO MEN HOLDING HANDS	Miscellaneous Sy...
💑	0x1f491	82	0.682683	5	16	61	COUPLE WITH HEART	Miscellaneous Sy...
🎃	0x1f383	81	0.723541	6	19	56	JACK-O-LANTERN	Miscellaneous Sy...
🐣	0x1f423	81	0.775651	12	17	52	HATCHING CHICK	Miscellaneous Sy...
▬	0x25ac	81	0.334743	0	42	39	BLACK RECTANGLE	Geometric Shapes
🔍	0xffff	81	0.308958	44	33	4	OBJECT REPLACEMENT CHA...	Specials
😟	0x1f61f	80	0.824309	28	18	34	WORRIED FACE	Emoticons
🐾	0x1f43e	78	0.738983	8	13	57	PAW PRINTS	Miscellaneous Sy...
🎓	0x1f393	77	0.647134	7	18	52	GRADUATION CAP	Miscellaneous Sy...
🏊	0x1f3ca	77	0.653096	2	27	48	SWIMMER	Miscellaneous Sy...
🍫	0x1f36b	76	0.792701	22	20	34	CHOCOLATE BAR	Miscellaneous Sy...
📷	0x1f4f7	76	0.489733	7	28	41	CAMERA	Miscellaneous Sy...
👄	0x1f444	75	0.591674	5	28	42	MOUTH	Miscellaneous Sy...
🌺	0x1f33c	74	0.603679	1	12	61	BLOSSOM	Miscellaneous Sy...
🚶	0x1f6b6	74	0.7969	28	29	17	PEDESTRIAN	Transport and Ma...
😺	0x1f431	73	0.544448	10	14	49	CAT FACE	Miscellaneous Sy...
▬	0x2551	73	0.568238	0	62	11	BOX DRAWINGS DOUBLE VE...	Box Drawing

	0x1f438	72	0.845772	24	30	18	FROG FACE	Miscellaneous Sy...
u	0x1f1fa	71	0.682893	8	15	48	REGIONAL INDICATOR SYM...	Enclosed Alphanu...
	0x1f47f	70	0.910731	47	15	8	IMP	Miscellaneous Sy...
	0x1f6ac	70	0.805066	8	16	46	SMOKING SYMBOL	Transport and Ma...
	0x273f	70	0.49871	2	38	30	BLACK FLORETTE	Dingbats
	0x1f30d	68	0.623864	5	16	47	EARTH GLOBE EUROPE-AFRI...	Miscellaneous Sy...
	0x1f412	68	0.639848	1	29	38	MONKEY	Miscellaneous Sy...
	0x1f4d6	68	0.709941	18	20	30	OPEN BOOK	Miscellaneous Sy...
...	0x250a	68	0.681885	0	0	68	BOX DRAWINGS LIGHT QUA...	Box Drawing
	0x1f425	67	0.779887	4	18	45	FRONT-FACING BABY CHICK	Miscellaneous Sy...
	0x1f300	66	0.553406	10	39	17	CYCLONE	Miscellaneous Sy...
	0x1f3a5	66	0.675876	8	30	28	MOVIE CAMERA	Miscellaneous Sy...
	0x1f43c	66	0.614422	4	40	22	PANDA FACE	Miscellaneous Sy...
i	0x1f484	66	0.636585	6	24	36	LIPSTICK	Miscellaneous Sy...
	0x1f4b8	66	0.790338	16	23	27	MONEY WITH WINGS	Miscellaneous Sy...
•	0x25cf	65	0.561224	8	37	20	BLACK CIRCLE	Geometric Shapes
	0x26d4	65	0.405987	5	22	38	NO ENTRY	Miscellaneous Sy...
	0x1f3c0	64	0.807197	13	21	30	BASKETBALL AND HOOP	Miscellaneous Sy...
	0x1f489	64	0.749556	11	18	35	SYRINGE	Miscellaneous Sy...
	0x1f49f	63	0.787621	3	12	48	HEART DECORATION	Miscellaneous Sy...
	0x1f4dd	62	0.794712	8	31	23	MEMO	Miscellaneous Sy...
	0x1f62f	62	0.859008	16	22	24	HUSHED FACE	Emoticons
	0x1f697	62	0.638134	8	31	23	AUTOMOBILE	Transport and Ma...
=	0x2550	62	0.441188	0	61	1	BOX DRAWINGS DOUBLE HO...	Box Drawing
	0x2666	61	0.562227	6	20	35	BLACK DIAMOND SUIT	Miscellaneous Sy...
	0x1f4ad	60	0.687955	10	27	23	THOUGHT BALLOON	Miscellaneous Sy...
☾	0x1f319	58	0.66939	3	16	39	CRESCENT MOON	Miscellaneous Sy...
	0x1f41f	58	0.676574	2	12	44	FISH	Miscellaneous Sy...
	0x1f463	58	0.638779	6	25	27	FOOTPRINTS	Miscellaneous Sy...
	0x1f5ff	58	0.941937	5	21	32	MOYAI	Miscellaneous Sy...
☞	0x261e	58	0.725402	0	51	7	WHITE RIGHT POINTING IN...	Miscellaneous Sy...
	0x2702	58	0.810272	34	18	6	BLACK SCISSORS	Dingbats
	0x1f35d	57	0.564422	16	18	23	SPAGHETTI	Miscellaneous Sy...
	0x1f36d	57	0.632539	9	21	27	LOLLIPOP	Miscellaneous Sy...
	0x1f46a	57	0.653304	24	10	23	FAMILY	Miscellaneous Sy...
	0x1f303	56	0.747624	3	27	26	NIGHT WITH STARS	Miscellaneous Sy...
X	0x274c	56	0.612998	14	12	30	CROSS MARK	Dingbats
	0x1f430	55	0.707517	4	13	38	RABBIT FACE	Miscellaneous Sy...
	0x1f48a	55	0.675071	8	14	33	PILL	Miscellaneous Sy...
	0x1f6a8	55	0.520959	6	6	43	POLICE CARS REVOLVING LI...	Transport and Ma...
	0x1f36a	54	0.736575	10	16	28	COOKIE	Miscellaneous Sy...
	0x1f626	54	0.882219	30	15	9	FROWNING FACE WITH OPE...	Emoticons
	0x1f363	53	0.66997	24	18	11	SUSHI	Miscellaneous Sy...
r	0x256d	53	0.490077	3	38	12	BOX DRAWINGS LIGHT ARC ...	Box Drawing
	0x2727	53	0.398521	1	33	19	WHITE FOUR POINTED STAR	Dingbats
	0x1f386	52	0.79852	2	9	41	FIREWORKS	Miscellaneous Sy...
r	0x256e	52	0.422827	5	35	12	BOX DRAWINGS LIGHT ARC ...	Box Drawing
◁	0x1f1e9	51	0.745012	1	16	34	REGIONAL INDICATOR SYM...	Enclosed Alphanu...
	0x1f38e	51	0.417135	0	2	49	JAPANESE DOLLS	Miscellaneous Sy...
	0x2705	51	0.672789	3	23	25	WHITE HEAVY CHECK MARK	Dingbats
	0x1f351	49	0.739306	12	12	25	PEACH	Miscellaneous Sy...
	0x1f3bc	49	0.636142	2	28	19	MUSICAL SCORE	Miscellaneous Sy...
	0x1f479	49	0.701173	13	20	16	JAPANESE OGRE	Miscellaneous Sy...
	0x1f4f1	49	0.719713	4	25	20	MOBILE PHONE	Miscellaneous Sy...
	0x1f50a	49	0.651739	4	20	25	SPEAKER WITH THREE SOU...	Miscellaneous Sy...
	0x1f64d	49	0.844923	29	8	12	PERSON FROWNING	Emoticons
	0x1f30c	47	0.750586	5	11	31	MILKY WAY	Miscellaneous Sy...
	0x1f34e	47	0.750898	6	19	22	RED APPLE	Miscellaneous Sy...
	0x1f43b	47	0.691609	6	13	28	BEAR FACE	Miscellaneous Sy...
—	0x2500	47	0.711348	0	40	7	BOX DRAWINGS LIGHT HORI...	Box Drawing
⤵	0x2570	47	0.509746	7	36	4	BOX DRAWINGS LIGHT ARC ...	Box Drawing
	0x1f487	46	0.678429	7	16	23	HAIRCUT	Miscellaneous Sy...
	0x265a	46	0.389195	0	44	2	BLACK CHESS KING	Miscellaneous Sy...
	0x266c	46	0.677442	4	26	16	BEAMED SIXTEENTH NOTES	Miscellaneous Sy...
	0x1f34a	45	0.6973	6	13	26	TANGERINE	Miscellaneous Sy...

	0x1f352	45	0.619434	6	18	21	CHERRIES	Miscellaneous Sy...
	0x1f371	45	0.58001	25	10	10	BENTO BOX	Miscellaneous Sy...
	0x1f42d	45	0.822546	0	12	33	MOUSE FACE	Miscellaneous Sy...
	0x1f45f	45	0.561643	5	15	25	ATHLETIC SHOE	Miscellaneous Sy...
	0x1f534	45	0.655928	1	24	20	LARGE RED CIRCLE	Miscellaneous Sy...
	0x1f30e	44	0.743439	5	19	20	EARTH GLOBE AMERICAS	Miscellaneous Sy...
	0x1f34d	44	0.583336	7	8	29	PINEAPPLE	Miscellaneous Sy...
	0x1f42e	43	0.846659	2	12	29	COW FACE	Miscellaneous Sy...
	0x1f4f2	43	0.598172	7	18	18	MOBILE PHONE WITH RIGHT...	Miscellaneous Sy...
	0x263c	43	0.461143	2	30	11	WHITE SUN WITH RAYS	Miscellaneous Sy...
	0x1f1f7	42	0.716328	1	10	31	REGIONAL INDICATOR SYM...	Enclosed Alphanu...
	0x1f305	42	0.603551	6	14	22	SUNRISE	Miscellaneous Sy...
	0x1f33d	42	0.648931	5	12	25	EAR OF MAIZE	Miscellaneous Sy...
	0x1f460	42	0.637243	6	14	22	HIGH-HEELED SHOE	Miscellaneous Sy...
	0x1f36c	41	0.784072	8	9	24	CANDY	Miscellaneous Sy...
	0x1f4a7	41	0.767648	20	8	13	DROPLET	Miscellaneous Sy...
	0x2753	41	0.614921	7	24	10	BLACK QUESTION MARK OR...	Dingbats
	0x1f35c	40	0.577917	9	5	26	STEAMING BOWL	Miscellaneous Sy...
	0x1f367	40	0.674848	9	9	22	SHAVED ICE	Miscellaneous Sy...
	0x1f3ac	40	0.650029	2	24	14	CLAPPER BOARD	Miscellaneous Sy...
	0x1f40f	40	0.975592	2	12	26	RAM	Miscellaneous Sy...
	0x1f418	40	0.722257	9	21	10	ELEPHANT	Miscellaneous Sy...
	0x1f434	40	0.795776	8	21	11	HORSE FACE	Miscellaneous Sy...
	0x1f467	40	0.649499	10	14	16	GIRL	Miscellaneous Sy...
	0x1f4a2	40	0.560644	6	18	16	ANGER SYMBOL	Miscellaneous Sy...
	0x1f63a	40	0.825898	8	7	25	SMILING CAT FACE WITH O...	Emoticons
	0x1f680	40	0.69848	2	15	23	ROCKET	Transport and Ma...
	0x2800	40	0.377951	6	26	8	BRAILLE PATTERN BLANK	Braille Patterns
	0xa6	40	0.402048	7	1	32	BROKEN BAR	Latin-1 Supplement
	0x1f3c4	39	0.772893	2	13	24	SURFER	Miscellaneous Sy...
	0x27a4	39	0.382195	0	26	13	BLACK RIGHTWARDS ARRO...	Dingbats
	0x2b06	39	0.605682	1	25	13	UPWARDS BLACK ARROW	Miscellaneous Sy...
	0x1f197	38	0.852317	6	4	28	SQUARED OK	Enclosed Alphanu...
	0x1f34b	38	0.742803	6	16	16	LEMON	Miscellaneous Sy...
	0x1f345	37	0.838127	9	5	23	TOMATO	Miscellaneous Sy...
	0x1f422	37	0.621647	6	17	14	TURTLE	Miscellaneous Sy...
	0x1f459	37	0.697202	3	11	23	BIKINI	Miscellaneous Sy...
	0x1f4fa	37	0.613869	3	16	18	TELEVISION	Miscellaneous Sy...
	0x26aa	37	0.56478	0	19	18	MEDIUM WHITE CIRCLE	Miscellaneous Sy...
	0x26c5	37	0.616324	4	11	22	SUN BEHIND CLOUD	Miscellaneous Sy...
	0x1f33e	36	0.558281	5	5	26	EAR OF RICE	Miscellaneous Sy...
	0x1f3e1	36	0.744479	5	9	22	HOUSE WITH GARDEN	Miscellaneous Sy...
	0x25c9	36	0.0466977	0	26	10	FISHEYE	Geometric Shapes
	0x1f1f9	35	0.584325	0	13	22	REGIONAL INDICATOR SYM...	Enclosed Alphanu...
	0x1f364	35	0.513142	6	21	8	FRIED SHRIMP	Miscellaneous Sy...
	0x1f41d	35	0.734687	5	17	13	HONEYBEE	Miscellaneous Sy...
	0x1f42c	35	0.746643	2	15	18	DOLPHIN	Miscellaneous Sy...
	0x2663	35	0.666097	4	14	17	BLACK CLUB SUIT	Miscellaneous Sy...
	0x270f	35	0.6251	3	16	16	PENCIL	Dingbats
	0x1f1ee	34	0.575926	0	12	22	REGIONAL INDICATOR SYM...	Enclosed Alphanu...
	0x1f31d	34	0.718125	5	17	12	FULL MOON WITH FACE	Miscellaneous Sy...
	0x1f40d	34	0.622535	3	15	16	SNAKE	Miscellaneous Sy...
	0x1f50b	34	0.583005	22	8	4	BATTERY	Miscellaneous Sy...
	0x2654	34	0.588904	0	14	20	WHITE CHESS KING	Miscellaneous Sy...
	0x1f315	33	0.641922	3	7	23	FULL MOON SYMBOL	Miscellaneous Sy...
	0x1f373	33	0.798473	13	6	14	COOKING	Miscellaneous Sy...
	0x1f428	33	0.696298	4	9	20	KOALA	Miscellaneous Sy...
	0x1f4bf	33	0.444466	0	11	22	OPTICAL DISC	Miscellaneous Sy...
	0x1f510	33	0.845214	6	9	18	CLOSED LOCK WITH KEY	Miscellaneous Sy...
	0x1f535	33	0.680544	4	14	15	LARGE BLUE CIRCLE	Miscellaneous Sy...
	0x1f63e	33	0.914645	20	5	8	POUTING CAT FACE	Emoticons
	0x2741	33	0.492536	0	31	2	EIGHT PETALLED OUTLINED ...	Dingbats
	0x1f333	32	0.53618	2	11	19	DECIDUOUS TREE	Miscellaneous Sy...
	0x1f470	32	0.787392	3	9	20	BRIDE WITH VEIL	Miscellaneous Sy...
	0x1f6b4	32	0.573442	0	9	23	BICYCLIST	Transport and Ma...

■	0x2580	32	0.249971	10	17	5	UPPER HALF BLOCK	Block Elements
⚓	0x2693	32	0.592873	2	8	22	ANCHOR	Miscellaneous Sy...
☼	0x2740	32	0.443664	0	18	14	WHITE FLORETTE	Dingbats
👗	0x1f457	31	0.70306	5	13	13	DRESS	Miscellaneous Sy...
⚡	0x2795	31	0.0341113	4	5	22	HEAVY PLUS SIGN	Dingbats
🍦	0x1f368	30	0.656537	8	7	15	ICE CREAM	Miscellaneous Sy...
🎈	0x1f4ac	30	0.635323	4	10	16	SPEECH BALLOON	Miscellaneous Sy...
💵	0x1f4b2	30	0.793676	0	22	8	HEAVY DOLLAR SIGN	Miscellaneous Sy...
👉	0x1f51c	30	0.606237	5	11	14	SOON WITH RIGHTWARDS ...	Miscellaneous Sy...
☯	0x2592	30	0.500646	2	27	1	MEDIUM SHADE	Block Elements
🚰	0x26fd	30	0.28735	1	23	6	FUEL PUMP	Miscellaneous Sy...
🍗	0x1f357	29	0.674706	9	9	11	POULTRY LEG	Miscellaneous Sy...
🍚	0x1f359	29	0.566487	5	10	14	RICE BALL	Miscellaneous Sy...
🐟	0x1f365	29	0.579634	22	4	3	FISH CAKE WITH SWIRL DES...	Miscellaneous Sy...
🍲	0x1f372	29	0.672925	7	11	11	POT OF FOOD	Miscellaneous Sy...
▶	0x25b8	29	0.516845	0	22	7	BLACK RIGHT-POINTING SM...	Geometric Shapes
♚	0x265b	29	0.489959	4	15	10	BLACK CHESS QUEEN	Miscellaneous Sy...
🍖	0x1f356	28	0.678554	7	10	11	MEAT ON BONE	Miscellaneous Sy...
🍛	0x1f35a	28	0.636903	6	2	20	COOKED RICE	Miscellaneous Sy...
🎹	0x1f3b9	28	0.466338	0	17	11	MUSICAL KEYBOARD	Miscellaneous Sy...
🐙	0x1f419	28	0.677638	2	12	14	OCTOPUS	Miscellaneous Sy...
👤	0x1f468	28	0.693755	2	8	18	MAN	Miscellaneous Sy...
🐱	0x1f63c	28	0.855991	5	7	16	CAT FACE WITH WRY SMILE	Emoticons
☯	0x2583	28	0.373188	0	0	28	LOWER THREE EIGHTHS BLO...	Block Elements
♚	0x2655	28	0.62131	0	16	12	WHITE CHESS QUEEN	Miscellaneous Sy...
♨	0x2668	28	0.664942	0	1	27	HOT SPRINGS	Miscellaneous Sy...
Ⓐ	0x1f1e7	27	0.595071	2	15	10	REGIONAL INDICATOR SYM...	Enclosed Alphanu...
Ⓒ	0x1f1ec	27	0.582061	2	15	10	REGIONAL INDICATOR SYM...	Enclosed Alphanu...
🍏	0x1f34f	27	0.852432	7	11	9	GREEN APPLE	Miscellaneous Sy...
👦	0x1f466	27	0.606458	5	13	9	BOY	Miscellaneous Sy...
👧	0x1f469	27	0.731773	9	7	11	WOMAN	Miscellaneous Sy...
🚗	0x1f698	27	0.649503	8	9	10	ONCOMING AUTOMOBILE	Transport and Ma...
☠	0x2620	27	0.748862	7	14	6	SKULL AND CROSSBONES	Miscellaneous Sy...
🐠	0x1f420	26	0.767681	4	6	16	TROPICAL FISH	Miscellaneous Sy...
🏠	0x1f4b5	26	0.806506	3	9	14	BANKNOTE WITH DOLLAR SI...	Miscellaneous Sy...
♂	0x1f6b9	26	0.811647	1	4	21	MENS SYMBOL	Transport and Ma...
📦	0x2560	26	0.542512	0	20	6	BOX DRAWINGS DOUBLE VE...	Box Drawing
☆	0x2730	26	0.194812	0	3	23	SHADOWED WHITE STAR	Dingbats
🌐	0x1f30f	25	0.68146	3	10	12	EARTH GLOBE ASIA-AUSTRA...	Miscellaneous Sy...
🌱	0x1f331	25	0.561338	0	9	16	SEEDLING	Miscellaneous Sy...
👛	0x1f45b	25	0.69636	4	7	14	PURSE	Miscellaneous Sy...
💻	0x1f4bb	25	0.489416	1	16	8	PERSONAL COMPUTER	Miscellaneous Sy...
🚙	0x1f699	25	0.653234	5	14	6	RECREATIONAL VEHICLE	Transport and Ma...
■	0x2584	25	0.464454	8	11	6	LOWER HALF BLOCK	Block Elements
👓	0x1f453	24	0.514452	0	16	8	EYEGLASSES	Miscellaneous Sy...
◀	0x25c4	24	0.671979	0	18	6	BLACK LEFT-POINTING POIN...	Geometric Shapes
⚾	0x26be	24	0.21361	9	7	8	BASEBALL	Miscellaneous Sy...
🌲	0x1f332	23	0.496443	2	9	12	EVERGREEN TREE	Miscellaneous Sy...
🌵	0x1f335	23	0.67006	6	6	11	CACTUS	Miscellaneous Sy...
🍇	0x1f347	23	0.583781	5	6	12	GRAPES	Miscellaneous Sy...
🍪	0x1f358	23	0.620778	5	3	15	RICE CRACKER	Miscellaneous Sy...
🍛	0x1f35b	23	0.585489	8	6	9	CURRY AND RICE	Miscellaneous Sy...
🏠	0x1f3e0	23	0.641778	1	8	14	HOUSE BUILDING	Miscellaneous Sy...
🐰	0x1f407	23	0.967358	4	9	10	RABBIT	Miscellaneous Sy...
👴	0x1f474	23	0.783858	5	7	11	OLDER MAN	Miscellaneous Sy...
👵	0x1f475	23	0.818007	5	2	16	OLDER WOMAN	Miscellaneous Sv...
👤	0x1f519	23	0.143849	2	14	7	BACK WITH LEFTWARDS AR...	Miscellaneous Sy...
🚫	0x1f51e	23	0.814491	6	12	5	NO ONE UNDER EIGHTEEN S...	Miscellaneous Sy...
◀	0x25c0	23	0.221018	0	16	7	BLACK LEFT-POINTING TRIA...	Geometric Shapes
🍮	0x1f36e	22	0.701403	10	5	7	CUSTARD	Miscellaneous Sy...
🎆	0x1f387	22	0.820989	1	3	18	FIREWORK SPARKLER	Miscellaneous Sy...
🐎	0x1f40e	22	0.698088	2	8	12	HORSE	Miscellaneous Sy...
🐣	0x1f424	22	0.838037	1	7	14	BABY CHICK	Miscellaneous Sy...
🐷	0x1f43d	22	0.647827	5	9	8	PIG NOSE	Miscellaneous Sy...
🏠	0x1f4b6	22	0.74829	3	16	3	BANKNOTE WITH EURO SIGN	Miscellaneous Sy...

	0x2569	22	0.811364	0	17	5	BOX DRAWINGS DOUBLE UP ...	Box Drawing
	0x2794	22	0.540919	3	19	0	HEAVY WIDE-HEADED RIGH...	Dingbats
	0x1f239	21	0.349908	1	12	8	SQUARED CJK UNIFIED IDE...	Enclosed Ideogra...
	0x1f311	21	0.603916	2	6	13	NEW MOON SYMBOL	Miscellaneous Sy...
	0x1f35e	21	0.710838	6	8	7	BREAD	Miscellaneous Sy...
	0x1f3be	21	0.820393	2	4	15	TENNIS RACQUET AND BALL	Miscellaneous Sy...
	0x1f3c1	21	0.618315	0	9	12	CHEQUERED FLAG	Miscellaneous Sy...
	0x1f411	21	0.921936	6	13	2	SHEEP	Miscellaneous Sy...
	0x1f6b2	21	0.582675	1	10	10	BICYCLE	Transport and Ma...
	0x1f6c0	21	0.904298	6	6	9	BATH	Transport and Ma...
	0x255a	21	0.714312	0	14	7	BOX DRAWINGS DOUBLE UP ...	Box Drawing
	0x1f433	20	0.655719	4	9	7	SPOUTING WHALE	Miscellaneous Sy...
	0x1f435	20	0.630508	1	7	12	MONKEY FACE	Miscellaneous Sy...
	0x1f46e	20	0.742231	10	8	2	POLICE OFFICER	Miscellaneous Sy...
	0x1f5fc	20	0.653789	1	6	13	TOKYO TOWER	Miscellaneous Sy...
	0x2533	20	0.399288	10	8	2	BOX DRAWINGS HEAVY DO...	Box Drawing
	0x2590	20	0.304139	4	15	1	RIGHT HALF BLOCK	Block Elements
	0x25d5	20	0.594134	0	10	10	CIRCLE WITH ALL BUT UPPE...	Geometric Shapes
	0x2639	20	0.77656	14	4	2	WHITE FROWNING FACE	Miscellaneous Sy...
	0x2660	20	0.533121	0	13	7	BLACK SPADE SUIT	Miscellaneous Sy...
	0x272a	20	0.287582	0	13	7	CIRCLED WHITE STAR	Dingbats
	0x1f192	19	0.627794	0	11	8	SQUARED COOL	Enclosed Alphanu...
	0x1f313	19	0.458297	0	6	13	FIRST QUARTER MOON SYM...	Miscellaneous Sy...
	0x1f36f	19	0.779823	7	4	8	HONEY POT	Miscellaneous Sy...
	0x1f375	19	0.650326	3	5	11	TEACUP WITHOUT HANDLE	Miscellaneous Sy...
	0x1f41a	19	0.526459	1	12	6	SPIRAL SHELL	Miscellaneous Sy...
	0x1f43a	19	0.734917	2	10	7	WOLF FACE	Miscellaneous Sy...
	0x1f442	19	0.682128	7	8	4	EAR	Miscellaneous Sy...
	0x1f47a	19	0.824825	8	7	4	JAPANESE GOBLIN	Miscellaneous Sy...
	0x1f512	19	0.939297	5	5	9	LOCK	Miscellaneous Sy...
	0x1f5fd	19	0.651088	1	10	8	STATUE OF LIBERTY	Miscellaneous Sy...
	0x21e8	19	0.541146	0	9	10	RIGHTWARDS WHITE ARROW	Arrows
	0x256c	19	0.321588	6	10	3	BOX DRAWINGS DOUBLE VE...	Box Drawing
	0x27a8	19	0.558287	2	12	5	HEAVY CONCAVE-POINTED ...	Dingbats
	0x1f302	18	0.784983	2	9	7	CLOSED UMBRELLA	Miscellaneous Sy...
	0x1f361	18	0.699734	5	9	4	DANGO	Miscellaneous Sy...
	0x1f3a1	18	0.506097	1	10	7	FERRIS WHEEL	Miscellaneous Sy...
	0x1f473	18	0.816081	2	1	15	MAN WITH TURBAN	Miscellaneous Sy...
	0x1f68c	18	0.635613	2	10	6	BUS	Transport and Ma...
	0x2669	18	0.791817	0	7	11	QUARTER NOTE	Miscellaneous Sy...
	0x2765	18	0.440217	0	13	5	ROTATED HEAVY BLACK HEA...	Dingbats
	0x1f31b	17	0.679967	0	6	11	FIRST QUARTER MOON WIT...	Miscellaneous Sy...
	0x1f31c	17	0.604117	0	7	10	LAST QUARTER MOON WITH...	Miscellaneous Sy...
	0x1f414	17	0.627149	1	9	7	CHICKEN	Miscellaneous Sy...
	0x1f416	17	0.814055	0	14	3	PIG	Miscellaneous Sy...
	0x1f429	17	0.870342	0	10	7	POODLE	Miscellaneous Sy...
	0x1f482	17	0.691283	8	4	5	GUARDSMAN	Miscellaneous Sy...
	0x1f48c	17	0.805001	2	3	12	LOVE LETTER	Miscellaneous Sy...
	0x1f506	17	0.604618	0	6	11	HIGH BRIGHTNESS SYMBOL	Miscellaneous Sy...
	0x1f64e	17	0.640743	8	2	7	PERSON WITH POUTING FACE	Emoticons
	0x1f6bf	17	0.669217	1	3	13	SHOWER	Transport and Ma...
	0x231a	17	0.649298	2	9	6	WATCH	Miscellaneous Tec...
	0x1f1eb	16	0.718034	0	7	9	REGIONAL INDICATOR SYM...	Enclosed Alphanu...
	0x1f314	16	0.447188	0	5	11	WAXING GIBBOUS MOON SY...	Miscellaneous Sy...
	0x1f318	16	0.577817	0	5	11	WANING CRESCENT MOON ...	Miscellaneous Sy...
	0x1f350	16	0.844576	5	3	8	PEAR	Miscellaneous Sy...
	0x1f3e9	16	0.587431	2	4	10	LOVE HOTEL	Miscellaneous Sy...
	0x1f426	16	0.757536	0	8	8	BIRD	Miscellaneous Sy...
	0x1f432	16	0.457546	2	13	1	DRAGON FACE	Miscellaneous Sy...
	0x1f4e2	16	0.545722	0	8	8	PUBLIC ADDRESS LOUDSPEA...	Miscellaneous Sy...
	0x1f528	16	0.738604	6	6	4	HAMMER	Miscellaneous Sy...
	0x2565	16	0.869483	6	2	8	BOX DRAWINGS DOWN DOU...	Box Drawing
	0x267b	16	0.438691	0	7	9	BLACK UNIVERSAL RECYCLI...	Miscellaneous Sy...
	0x274a	16	0.452128	0	16	0	EIGHT TEARDROP-SPOKED P...	Dingbats
	0x1f1f4	15	0.676124	0	9	6	REGIONAL INDICATOR SYM...	Enclosed Alphanu...

☾	0x1f316	15	0.534539	0	4	11	WANING GIBBOUS MOON SY...	Miscellaneous Sy...
☾	0x1f317	15	0.54682	0	4	11	LAST QUARTER MOON SYMBOL	Miscellaneous Sy...
☾	0x1f362	15	0.684276	5	7	3	ODEN	Miscellaneous Sy...
🎨	0x1f3a8	15	0.490548	1	10	4	ARTIST PALETTE	Miscellaneous Sy...
🎭	0x1f3ad	15	0.65766	5	5	5	PERFORMING ARTS	Miscellaneous Sy...
🐮	0x1f404	15	0.825888	2	6	7	COW	Miscellaneous Sy...
👖	0x1f456	15	0.910195	1	7	7	JEANS	Miscellaneous Sy...
☼	0x1f505	15	0.634783	0	0	15	LOW BRIGHTNESS SYMBOL	Miscellaneous Sy...
😘	0x1f617	15	0.714328	1	2	12	KISSING FACE	Emoticons
🚶	0x1f6ba	15	0.677227	2	8	5	WOMENS SYMBOL	Transport and Ma...
👶	0x1f6bc	15	0.692875	0	5	10	BABY SYMBOL	Transport and Ma...
—	0x2581	15	0.71132	0	15	0	LOWER ONE EIGHTH BLOCK	Block Elements
↙	0x25df	15	0.646546	2	12	1	LOWER LEFT QUADRANT CI...	Geometric Shapes
🗿	0x26f2	15	0.69731	1	12	2	FOUNTAIN	Miscellaneous Sy...
⬇️	0x2b07	15	0.631917	2	4	9	DOWNWARDS BLACK ARROW	Miscellaneous Sy...
🌫️	0x1f301	14	0.744407	1	9	4	FOGGY	Miscellaneous Sy...
★	0x1f320	14	0.761412	2	1	11	SHOOTING STAR	Miscellaneous Sy...
🔔	0x1f390	14	0.29088	0	11	3	WIND CHIME	Miscellaneous Sy...
🎢	0x1f3a2	14	0.825804	1	4	9	ROLLER COASTER	Miscellaneous Sy...
🎷	0x1f3b7	14	0.759955	0	3	11	SAXOPHONE	Miscellaneous Sy...
🏰	0x1f3f0	14	0.434067	1	7	6	EUROPEAN CASTLE	Miscellaneous Sy...
🐱	0x1f408	14	0.645414	0	9	5	CAT	Miscellaneous Sy...
🐌	0x1f40c	14	0.42135	0	3	11	SNAIL	Miscellaneous Sy...
🐹	0x1f439	14	0.743701	3	3	8	HAMSTER FACE	Miscellaneous Sy...
👜	0x1f45c	14	0.675483	2	6	6	HANDBAG	Miscellaneous Sy...
👛	0x1f4bc	14	0.573028	1	3	10	BRIEFCASE	Miscellaneous Sy...
🚕	0x1f695	14	0.526017	3	8	3	TAXI	Transport and Ma...
🚴	0x1f6b5	14	0.556155	0	8	6	MOUNTAIN BICYCLIST	Transport and Ma...
⬆️	0x21e7	14	0.46142	0	12	2	UPWARDS WHITE ARROW	Arrows
—	0x2508	14	0.79274	12	0	2	BOX DRAWINGS LIGHT QUA...	Box Drawing
▢	0x2557	14	0.36112	0	8	6	BOX DRAWINGS DOUBLE DO...	Box Drawing
▧	0x2571	14	0.547959	0	14	0	BOX DRAWINGS LIGHT DIAG...	Box Drawing
☎️	0x260e	14	0.836709	2	10	2	BLACK TELEPHONE	Miscellaneous Sy...
♣️	0x2667	14	0.775062	0	6	8	WHITE CLUB SUIT	Miscellaneous Sy...
●	0x26ab	14	0.440913	3	4	7	MEDIUM BLACK CIRCLE	Miscellaneous Sy...
▣	0x1f194	13	0.479125	0	1	12	SQUARED ID	Enclosed Alphanu...
🌇	0x1f307	13	0.542874	0	5	8	SUNSET OVER BUILDINGS	Miscellaneous Sy...
🌑	0x1f312	13	0.446222	0	4	9	WAXING CRESCENT MOON S...	Miscellaneous Sy...
🛷	0x1f3bf	13	0.727392	1	5	7	SKI AND SKI BOOT	Miscellaneous Sy...
🐪	0x1f42a	13	0.836925	0	4	9	DROMEDARY CAMEL	Miscellaneous Sy...
🚂	0x1f682	13	0.656031	0	9	4	STEAM LOCOMOTIVE	Transport and Ma...
⬇️	0x21e9	13	0.503573	0	13	0	DOWNWARDS WHITE ARROW	Arrows
🕒	0x23f0	13	0.64578	2	2	9	ALARM CLOCK	Miscellaneous Tec...
▢	0x2554	13	0.309824	0	9	4	BOX DRAWINGS DOUBLE DO...	Box Drawing
⬆️	0x255d	13	0.800038	0	6	7	BOX DRAWINGS DOUBLE UP ...	Box Drawing
⤴️	0x25e0	13	0.652555	0	6	7	UPPER HALF CIRCLE	Geometric Shapes
🏪	0x26ea	13	0.387239	2	7	4	CHURCH	Miscellaneous Sy...
✦	0x2726	13	0.40237	0	12	1	BLACK FOUR POINTED STAR	Dingbats
▣	0x1f193	12	0.551459	0	9	3	SQUARED FREE	Enclosed Alphanu...
ℓ	0x1f1f1	12	0.713337	0	6	6	REGIONAL INDICATOR SYM...	Enclosed Alphanu...
🏙️	0x1f306	12	0.764797	2	6	4	CITYSCAPE AT DUSK	Miscellaneous Sy...
🍠	0x1f360	12	0.938678	5	2	5	ROASTED SWEET POTATO	Miscellaneous Sy...
🚪	0x1f3e7	12	0.88824	0	0	12	AUTOMATED TELLER MACHINE	Miscellaneous Sy...
🏪	0x1f3ea	12	0.374437	2	7	3	CONVENIENCE STORE	Miscellaneous Sy...
🐳	0x1f40b	12	0.865813	2	5	5	WHALE	Miscellaneous Sv...
🐛	0x1f41b	12	0.672752	2	4	6	BUG	Miscellaneous Sy...
🐫	0x1f42b	12	0.770082	0	7	5	BACTRIAN CAMEL	Miscellaneous Sy...
👔	0x1f454	12	0.520012	2	3	7	NECKTIE	Miscellaneous Sy...
👕	0x1f455	12	0.795098	1	4	7	T-SHIRT	Miscellaneous Sy...
💡	0x1f4a1	12	0.597619	0	3	9	ELECTRIC LIGHT BULB	Miscellaneous Sy...
💳	0x1f4b3	12	0.833923	1	4	7	CREDIT CARD	Miscellaneous Sy...
☪️	0x1f531	12	0.790792	3	5	4	TRIDENT EMBLEM	Miscellaneous Sy...
💠	0x1f539	12	0.1702	2	6	4	SMALL BLUE DIAMOND	Miscellaneous Sy...
🚊	0x1f68b	12	0.317741	0	11	1	TRAM CAR	Transport and Ma...
—	0x2582	12	0.578512	0	12	0	LOWER ONE QUARTER BLOCK	Block Elements

▽	0x25bd	12	0.843553	1	4	7	WHITE DOWN-POINTING TR...	Geometric Shapes
←	0x2b05	12	0.662575	0	5	7	LEFTWARDS BLACK ARROW	Miscellaneous Sy...
ﻻ	0x6e9	12	0.464313	0	12	0	ARABIC PLACE OF SAJDAH	Arabic
☒	0x1f198	11	0.775132	4	2	5	SQUARED SOS	Enclosed Alphanu...
⚽	0x1f3c9	11	0.837043	0	4	7	RUGBY FOOTBALL	Miscellaneous Sy...
🐭	0x1f400	11	0.701471	1	7	3	RAT	Miscellaneous Sy...
🐮	0x1f402	11	0.774388	1	7	3	OX	Miscellaneous Sy...
👚	0x1f45a	11	0.810044	0	3	8	WOMANS CLOTHES	Miscellaneous Sy...
👞	0x1f45e	11	0.779457	1	3	7	MANS SHOE	Miscellaneous Sy...
📹	0x1f4f9	11	0.718347	1	3	7	VIDEO CAMERA	Miscellaneous Sy...
🗻	0x1f5fb	11	0.723456	0	3	8	MOUNT FUJI	Miscellaneous Sy...
🚌	0x1f68d	11	0.659559	4	2	5	ONCOMING BUS	Transport and Ma...
🚑	0x1f691	11	0.758088	3	4	4	AMBULANCE	Transport and Ma...
🚤	0x1f6a3	11	0.712556	0	3	8	ROWBOAT	Transport and Ma...
📦	0x2566	11	0.220263	0	6	5	BOX DRAWINGS DOUBLE DO...	Box Drawing
◻	0x25a1	11	0.673637	7	0	4	WHITE SQUARE	Geometric Shapes
✳	0x2733	11	0.372346	0	11	0	EIGHT SPOKED ASTERISK	Dingbats
🌰	0x1f330	10	0.786442	0	3	7	CHESTNUT	Miscellaneous Sy...
🎒	0x1f392	10	0.614006	1	2	7	SCHOOL SATCHEL	Miscellaneous Sy...
🏂	0x1f3c2	10	0.777003	1	3	6	SNOWBOARDER	Miscellaneous Sy...
🏢	0x1f3e2	10	0.542296	0	8	2	OFFICE BUILDING	Miscellaneous Sy...
🎓	0x1f3eb	10	0.64421	6	1	3	SCHOOL	Miscellaneous Sy...
🐉	0x1f409	10	0.628495	0	8	2	DRAGON	Miscellaneous Sy...
🐊	0x1f40a	10	0.845508	2	5	3	CROCODILE	Miscellaneous Sy...
🐕	0x1f415	10	0.879334	1	5	4	DOG	Miscellaneous Sy...
👡	0x1f461	10	0.63544	1	3	6	WOMANS SANDAL	Miscellaneous Sy...
📴	0x1f4f4	10	0.981899	1	0	9	MOBILE PHONE OFF	Miscellaneous Sy...
📻	0x1f4fb	10	0.7262	1	4	5	RADIO	Miscellaneous Sy...
🔔	0x1f514	10	0.713041	0	0	10	BELL	Miscellaneous Sy...
🚚	0x1f69a	10	0.479308	2	7	1	DELIVERY TRUCK	Transport and Ma...
🚢	0x1f6a2	10	0.817914	3	1	6	SHIP	Transport and Ma...
⤵	0x2312	10	0.623939	0	3	7	ARC	Miscellaneous Tec...
▲	0x25e2	10	0.571992	0	2	8	BLACK LOWER RIGHT TRIAN...	Geometric Shapes
⛖	0x26fa	10	0.905849	1	2	7	TENT	Miscellaneous Sy...
🖊	0x2712	10	0.543607	1	5	4	BLACK NIB	Dingbats
◻	0x2752	10	0.587413	0	7	3	UPPER RIGHT SHADOWED ...	Dingbats
♠	0x1f0cf	9	0.685607	3	2	4	PLAYING CARD BLACK JOKER	Playing Cards
🍈	0x1f348	9	0.520981	6	1	2	MELON	Miscellaneous Sy...
🎺	0x1f3ba	9	0.64618	1	0	8	TRUMPET	Miscellaneous Sy...
🎻	0x1f3bb	9	0.72088	0	5	4	VIOLIN	Miscellaneous Sy...
👕	0x1f3bd	9	0.903768	0	4	5	RUNNING SHIRT WITH SASH	Miscellaneous Sy...
🏥	0x1f3e5	9	0.648093	1	4	4	HOSPITAL	Miscellaneous Sy...
🐐	0x1f410	9	0.720583	0	4	5	GOAT	Miscellaneous Sy...
💒	0x1f492	9	0.794172	0	3	6	WEDDING	Miscellaneous Sy...
📅	0x1f4c5	9	0.597376	2	3	4	CALENDAR	Miscellaneous Sy...
🔒	0x1f513	9	0.80289	3	2	4	OPEN LOCK	Miscellaneous Sy...
↩	0x1f51a	9	0.868645	1	3	5	END WITH LEFTWARDS ARR...	Miscellaneous Sy...
💎	0x1f537	9	0.367845	0	7	2	LARGE BLUE DIAMOND	Miscellaneous Sy...
🚕	0x1f696	9	0.800145	3	4	2	ONCOMING TAXI	Transport and Ma...
🚤	0x1f6a4	9	0.503972	0	5	4	SPEEDBOAT	Transport and Ma...
🚦	0x1f6a6	9	0.454265	2	4	3	VERTICAL TRAFFIC LIGHT	Transport and Ma...
▬	0x2563	9	0.641758	0	9	0	BOX DRAWINGS DOUBLE VE...	Box Drawing
▬	0x258c	9	0.420712	4	4	1	LEFT HALF BLOCK	Block Elements
■	0x25a0	9	0.790444	5	2	2	BLACK SQUARE	Geometric Shapes
▼	0x25bc	9	0.656774	2	4	3	BLACK DOWN-POINTING TRI...	Geometric Shapes
▴	0x25e4	9	0.669224	0	0	9	BLACK UPPER LEFT TRIANGLE	Geometric Shapes
↗	0x261b	9	0.553783	0	5	4	BLACK RIGHT POINTING INDEX	Miscellaneous Sy...
✉	0x2709	9	0.44131	1	4	4	ENVELOPE	Dingbats
☆	0x2729	9	0.265668	0	9	0	STRESS OUTLINED WHITE S...	Dingbats
★	0x272f	9	0.320905	0	9	0	PINWHEEL STAR	Dingbats
?	0x2754	9	0.787467	3	3	3	WHITE QUESTION MARK OR...	Dingbats
🍄	0x1f344	8	0.481956	2	4	2	MUSHROOM	Miscellaneous Sy...
🍼	0x1f37c	8	0.889794	1	1	6	BABY BOTTLE	Miscellaneous Sy...
🐃	0x1f403	8	0.829644	0	8	0	WATER BUFFALO	Miscellaneous Sy...
🐘	0x1f417	8	0.756674	0	3	5	BOAR	Miscellaneous Sy...

	0x1f4b4	8	0.837158	3	4	1	BANKNOTE WITH YEN SIGN	Miscellaneous Sy...
	0x1f4c0	8	0.842925	1	5	2	DVD	Miscellaneous Sy...
	0x1f4d3	8	0.835156	1	4	3	NOTEBOOK	Miscellaneous Sy...
	0x1f4d5	8	0.687816	1	4	3	CLOSED BOOK	Miscellaneous Sy...
	0x1f4e3	8	0.665291	1	2	5	CHEERING MEGAPHONE	Miscellaneous Sy...
	0x1f50c	8	0.843838	3	3	2	ELECTRIC PLUG	Miscellaneous Sy...
	0x1f693	8	0.654503	0	5	3	POLICE CAR	Transport and Ma...
	0x1f69b	8	0.552176	2	6	0	ARTICULATED LORRY	Transport and Ma...
	0x21aa	8	0.726671	0	7	1	RIGHTWARDS ARROW WIT...	Arrows
	0x2503	8	0.495889	0	4	4	BOX DRAWINGS HEAVY VER...	Box Drawing
	0x251b	8	0.71163	0	2	6	BOX DRAWINGS HEAVY UP A...	Box Drawing
	0x253b	8	0.436071	6	0	2	BOX DRAWINGS HEAVY UP A...	Box Drawing
	0x253c	8	0.594639	2	4	2	BOX DRAWINGS LIGHT VERT...	Box Drawing
	0x25cb	8	0.625335	0	3	5	WHITE CIRCLE	Geometric Shapes
	0x2609	8	0.985925	0	6	2	SUN	Miscellaneous Sy...
	0x26f3	8	0.659687	0	1	7	FLAG IN HOLE	Miscellaneous Sy...
	0x27b0	8	0.255281	4	1	3	CURLY LOOP	Dingbats
	0x1f1f2	7	0.786774	0	3	4	REGIONAL INDICATOR SYM...	Enclosed Alphanu...
	0x1f309	7	0.848338	0	2	5	BRIDGE AT NIGHT	Miscellaneous Sy...
	0x1f3c7	7	0.829352	3	2	2	HORSE RACING	Miscellaneous Sy...
	0x1f3e8	7	0.803363	0	5	2	HOTEL	Miscellaneous Sy...
	0x1f406	7	0.550091	0	4	3	LEOPARD	Miscellaneous Sy...
	0x1f452	7	0.843649	1	3	3	WOMANS HAT	Miscellaneous Sy...
	0x1f471	7	0.584041	1	4	2	PERSON WITH BLOND HAIR	Miscellaneous Sy...
	0x1f4b7	7	0.573125	0	3	4	BANKNOTE WITH POUND SIGN	Miscellaneous Sy...
	0x1f4ba	7	0.499523	1	3	3	SEAT	Miscellaneous Sy...
	0x1f4c3	7	0.754189	0	3	4	PAGE WITH CURL	Miscellaneous Sy...
	0x1f4cc	7	0.799055	5	1	1	PUSHPIN	Miscellaneous Sy...
	0x1f4de	7	0.639664	0	3	4	TELEPHONE RECEIVER	Miscellaneous Sy...
	0x1f51b	7	0.474006	1	1	5	ON WITH EXCLAMATION MA...	Miscellaneous Sy...
	0x1f538	7	0.172105	0	5	2	SMALL ORANGE DIAMOND	Miscellaneous Sy...
	0x1f684	7	0.699814	0	2	5	HIGH-SPEED TRAIN	Transport and Ma...
	0x1f69c	7	0.767275	0	0	7	TRACTOR	Transport and Ma...
	0x1f6a9	7	0.146683	3	3	1	TRIANGULAR FLAG ON POST	Transport and Ma...
	0x231b	7	0.796788	2	2	3	HOURLASS	Miscellaneous Tec...
	0x23f3	7	0.829007	3	1	3	HOURLASS WITH FLOWIN...	Miscellaneous Tec...
	0x24b6	7	0.467453	2	4	1	CIRCLED LATIN CAPITAL LE...	Enclosed Alphanu...
	0x25b2	7	0.337837	0	2	5	BLACK UP-POINTING TRIAN...	Geometric Shapes
	0x263b	7	0.808079	1	3	3	BLACK SMILING FACE	Miscellaneous Sy...
	0x264e	7	0.85515	2	4	1	LIBRA	Miscellaneous Sy...
	0x2662	7	0.868165	0	4	3	WHITE DIAMOND SUIT	Miscellaneous Sy...
	0x26f5	7	0.323527	0	2	5	SAILBOAT	Miscellaneous Sy...
	0x270e	7	0.547315	0	2	5	LOWER RIGHT PENCIL	Dingbats
	0x2745	7	0.247154	0	1	6	TIGHT TRIFOLIATE SNOWFL...	Dingbats
	0x2755	7	0.862023	1	3	3	WHITE EXCLAMATION MARK...	Dingbats
	0x1f170	6	0.311139	1	2	3	NEGATIVE SQUARED LATIN ...	Enclosed Alphanu...
	0x1f1fc	6	0.734392	0	3	3	REGIONAL INDICATOR SYM...	Enclosed Alphanu...
	0x1f30b	6	0.37968	0	2	4	VOLCANO	Miscellaneous Sy...
	0x1f38b	6	0.785516	1	0	5	TANABATA TREE	Miscellaneous Sy...
	0x1f3aa	6	0.607881	0	4	2	CIRCUS TENT	Miscellaneous Sy...
	0x1f3ef	6	0.52992	0	5	1	JAPANESE CASTLE	Miscellaneous Sy...
	0x1f41c	6	0.879902	0	2	4	ANT	Miscellaneous Sy...
	0x1f421	6	0.867289	0	4	2	BLOWFISH	Miscellaneous Sy...
	0x1f45d	6	0.611228	1	1	4	POUCH	Miscellaneous Sy...
	0x1f465	6	0.837925	1	2	3	BUSTS IN SILHOUETTE	Miscellaneous Sy...
	0x1f477	6	0.739199	1	2	3	CONSTRUCTION WORKER	Miscellaneous Sy...
	0x1f4c4	6	0.62414	0	0	6	PAGE FACING UP	Miscellaneous Sy...
	0x1f4cd	6	0.488693	1	3	2	ROUND PUSHPIN	Miscellaneous Sy...
	0x1f4e1	6	0.553892	1	2	3	SATELLITE ANTENNA	Miscellaneous Sy...
	0x1f4e5	6	0.981884	0	6	0	INBOX TRAY	Miscellaneous Sy...
	0x1f4f5	6	0.794071	1	3	2	NO MOBILE PHONES	Miscellaneous Sy...
	0x1f508	6	0.635444	0	4	2	SPEAKER	Miscellaneous Sy...
	0x1f52d	6	0.715984	0	3	3	TELESCOPE	Miscellaneous Sy...
	0x1f690	6	0.776426	0	1	5	MINIBUS	Transport and Ma...
	0x1f694	6	0.521944	1	4	1	ONCOMING POLICE CAR	Transport and Ma...

🚫	0x1f6b3	6	0.102994	0	0	6	NO BICYCLES	Transport and Ma...
⬇	0x21b3	6	0.346429	0	6	0	DOWNWARDS ARROW WIT...	Arrows
➡	0x23e9	6	0.731714	0	5	1	BLACK RIGHT-POINTING DO...	Miscellaneous Tec...
⓪	0x24ba	6	0.814815	0	3	3	CIRCLED LATIN CAPITAL LE...	Enclosed Alphanu...
Ⓛ	0x24c1	6	0.726852	0	3	3	CIRCLED LATIN CAPITAL LE...	Enclosed Alphanu...
📦	0x2513	6	0.515225	0	2	4	BOX DRAWINGS HEAVY DO...	Box Drawing
📐	0x2523	6	0.487486	0	2	4	BOX DRAWINGS HEAVY VER...	Box Drawing
◆	0x25c6	6	0.602165	0	1	5	BLACK DIAMOND	Geometric Shapes
◡	0x25e1	6	0.946757	0	4	2	LOWER HALF CIRCLE	Geometric Shapes
☂	0x2602	6	0.45339	1	4	1	UMBRELLA	Miscellaneous Sy...
☑	0x2610	6	0.538095	6	0	0	BALLOT BOX	Miscellaneous Sy...
☾	0x263e	6	0.452908	0	6	0	LAST QUARTER MOON	Miscellaneous Sy...
♌	0x264c	6	0.79041	0	1	5	LEO	Miscellaneous Sy...
✕	0x2718	6	0.538603	0	1	5	HEAVY BALLOT X	Dingbats
†	0x271e	6	0.50939	0	3	3	SHADOWED WHITE LATIN C...	Dingbats
ﺀ	0x6de	6	0.471786	0	6	0	ARABIC START OF RUB EL H...	Arabic
⓪	0x1f17e	5	0.977469	2	2	1	NEGATIVE SQUARED LATIN ...	Enclosed Alphanu...
🌄	0x1f304	5	0.285584	0	4	1	SUNRISE OVER MOUNTAINS	Miscellaneous Sy...
🎠	0x1f3a0	5	0.574337	0	2	3	CAROUSEL HORSE	Miscellaneous Sy...
👒	0x1f3a9	5	0.586476	0	3	2	TOP HAT	Miscellaneous Sy...
📧	0x1f3e4	5	0.272576	0	3	2	EUROPEAN POST OFFICE	Miscellaneous Sy...
🐭	0x1f401	5	0.418556	2	3	0	MOUSE	Miscellaneous Sy...
🐯	0x1f405	5	0.81855	0	4	1	TIGER	Miscellaneous Sy...
👠	0x1f462	5	0.994118	0	1	4	WOMANS BOOTS	Miscellaneous Sy...
🗿	0x1f464	5	0.785821	0	1	4	BUST IN SILHOUETTE	Miscellaneous Sy...
📗	0x1f4d7	5	0.78916	1	2	2	GREEN BOOK	Miscellaneous Sy...
📯	0x1f4ef	5	0.974048	1	4	0	POSTAL HORN	Miscellaneous Sy...
📰	0x1f4f0	5	0.86015	1	1	3	NEWSPAPER	Miscellaneous Sy...
📡	0x1f4f6	5	0.930493	0	3	2	ANTENNA WITH BARS	Miscellaneous Sy...
🌀	0x1f504	5	0.971014	0	5	0	ANTICLOCKWISE DOWNWA...	Miscellaneous Sy...
🔦	0x1f526	5	0.852648	0	5	0	ELECTRIC TORCH	Miscellaneous Sy...
🔧	0x1f527	5	0.677226	4	0	1	WRENCH	Miscellaneous Sy...
🔩	0x1f529	5	0.672722	1	2	2	NUT AND BOLT	Miscellaneous Sy...
💎	0x1f536	5	0.786004	0	3	2	LARGE ORANGE DIAMOND	Miscellaneous Sy...
🗼	0x1f5fe	5	0.738586	0	3	2	SILHOUETTE OF JAPAN	Miscellaneous Sy...
🚁	0x1f681	5	0.509989	0	4	1	HELICOPTER	Transport and Ma...
🚦	0x1f6a5	5	0.723894	1	3	1	HORIZONTAL TRAFFIC LIGHT	Transport and Ma...
⓪	0x24c2	5	0.765244	1	1	3	CIRCLED LATIN CAPITAL LE...	Enclosed Alphanu...
📦	0x2510	5	0.770866	2	2	1	BOX DRAWINGS LIGHT DOW...	Box Drawing
◀	0x25c2	5	0.198501	0	3	2	BLACK LEFT-POINTING SMAL...	Geometric Shapes
↙	0x25de	5	0.722789	1	3	1	LOWER RIGHT QUADRANT C...	Geometric Shapes
☄	0x2604	5	0.435374	0	5	0	COMET	Miscellaneous Sy...
†	0x2628	5	0.457692	0	5	0	CROSS OF LORRAINE	Miscellaneous Sy...
☺	0x262e	5	0.750376	2	1	2	PEACE SYMBOL	Miscellaneous Sy...
♂	0x2642	5	0.580234	0	4	1	MALE SIGN	Miscellaneous Sy...
🎵	0x266e	5	0.936641	0	4	1	MUSIC NATURAL SIGN	Miscellaneous Sy...

แหล่งที่มา: http://kt.ijs.si/data/Emoji_sentiment_ranking/index.html

ประวัติผู้เขียน



นางสาวกนกกานต์ ด่านสว่างกุล

อีเมล kanokkant.kul@gmail.com

เบอร์โทรศัพท์ 09 2256 5708

การศึกษา สาขาวิชาวิทยาการคอมพิวเตอร์

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย