



UNIVERSITY OF  
BIRMINGHAM



# Emoji as Sentiment Indicators

An Investigative Case Study  
in Arabic Text

Shatha Hakami

School of / Dept. of Computer Science  
University of Birmingham, UK / Jazan University, Saudi Arabia  
sah624@cs.bham.ac.uk / sahakami@jazanu.edu.sa

Robert Hendley

School of Computer Science  
University of Birmingham, UK  
r.j.hendley@cs.bham.ac.uk

Phillip Smith

School of Computer Science  
University of Birmingham, UK  
p.smith.7@cs.bham.ac.uk



# Human Communication

## Face-to-Face

Verbal Cues: words and sentences.

Nonverbal Cues: eye contacts, head nodding, facial expressions, or vocal intonation.

## Text-Based

Verbal Cues: words and sentences.

Nonverbal Cues: punctuations, emoticons or emoji.



# Emoticons and Emoji

- **Emoticons** are a sequence of keyboard characters (ASCII characters).

(^\_^) (\*\_\*) (@\_@)

- **Emoji** are actual icons (Unicode characters) that appear on physical or virtual keyboards and can be used across various platforms.



- **Sentiment Analysis** is a process that analyses text and builds an interpretation of the sentiment that it is intended to convey.



# Related Work

- Emoji as Nonverbal Cues:
  - Emoji are visible acts of meaning.
- Emoji for Sentiment Analysis:
  - Facial and non-facial emoji.
  - Emoji as reliable ground truth for the sentiment.

## Limitation

- The conventional approach of performing separate sentiment analysis of text and emoji and then combining the two to generate an overall value, is **inadequate**.



RQ1. When is it appropriate, in sentiment analysis, to use the conventional techniques for interpreting emoji?

RQ2. What are the other, unconventional, cases of emoji in sentiment analysis, and when do they apply?

# Our Focus



# Methodology

Sentiment Analysis

+






Coding and Counting

- Observe (Datasets Collection)
  - Collect 5402 Arabic tweets with 496 unique emoji from seven already-existing datasets.
  - Create three datasets from the collected tweets:
    1. Emoji-Text dataset.
    2. Emoji-Only dataset.
    3. Plain-Text dataset.
- Code (Sentiment Annotation)
  - Unify all the labels to be in the sentiment-label form.
  - Human Annotation (two females, one male).
    - Inter-rater Fleiss' Kappa agreement test  $k = 0.85$ .
  - Machine Annotation (Mazajak Python model).
- Count (Frequency Analysis)
  - Compute the Frequency (Fr.) and Relative Frequency (RelFr.) of all the three datasets.
  - Counting Emoji Load.



# Results

TABLE I. THE TOP 5 EMOJI IN EMOJI-ONLY DATASET WITH SENTIMENT FREQUENCY (Fr.) AND RELATIVE FREQUENCY (RelFr.).

Emojis	Name	Class	Sentiment	Total	W/ Negative Texts Fr.(RelFr.)	W/ Neutral Texts Fr.(RelFr.)	W/ Positive Texts Fr.(RelFr.)
	Face with Tears of Joy	Facial Expression	Positive	2,270	1,229 (54.14%)	92 (4.05%)	949 (41.80%)
	Red Heart	Heart	Positive	765	45 (5.88%)	20 (2.61%)	700 (91.50%)
	Saudi Arabia	Flag	Positive	733	89 (12.14%)	29 (3.95%)	615 (83.90%)
	Smiling Face with Heart-Eyes	Facial Expression	Positive	426	21 (4.93%)	15 (3.52%)	390 (91.55%)
	Broken Heart	Heart	Negative	410	286 (69.75%)	16 (3.90%)	108 (26.34%)



# Results

TABLE II. THE FREQUENCY (Fr.) AND RELATIVE FREQUENCY (RelFr.) OF SENTIMENTS IN THE PLAIN-TEXT, EMOJI-TEXT AND EMOJI-ONLY DATASETS.

Sentiment Label	Plain-text Fr.(RelFr.)	Emoji-text Fr.(RelFr.)	Emoji-only Fr.(RelFr.)
Negative	2045 (39%)	1885 (36%)	4016 (31%)
Neutral	1119 (22%)	965 (19%)	2547 (20%)
Positive	2040 (39%)	2354 (45%)	6244 (49%)
Total	5,204	5,204	12,807

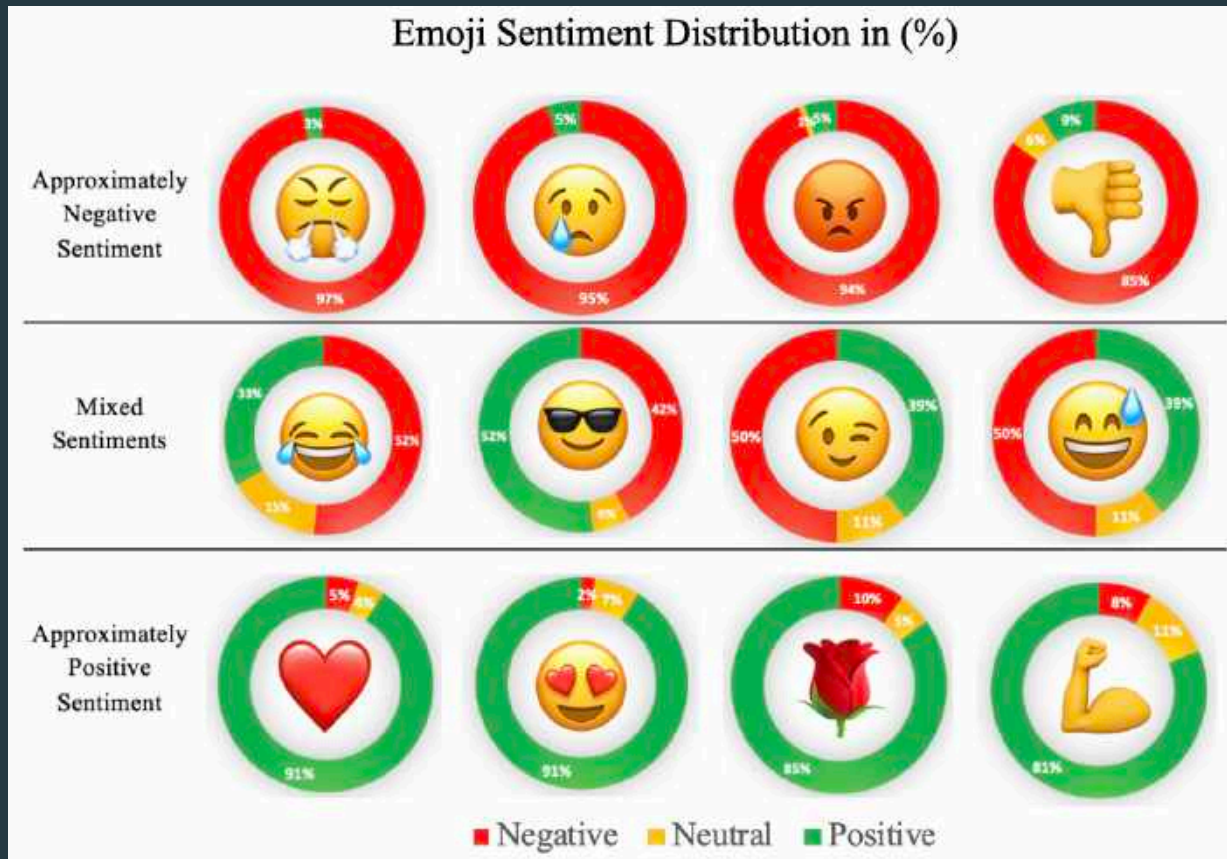
TABLE III. THE FREQUENCY (Fr.) AND RELATIVE FREQUENCY (RelFr.) OF SENTIMENTS IN THE EMOJI-TEXT DATASET WITH DIFFERENT EMOJI LOAD.

Emoji Load	Total Text Fr.(RelFr.)	Neg. Text Fr.(RelFr.)	Neut. Text Fr.(RelFr.)	Pos. Text Fr.(RelFr.)
1	2283 (44%)	908 (40%)	436 (19%)	939 (41%)
2	1358 (26%)	467 (34%)	233 (17%)	658 (48%)
3	652 (12%)	261 (40%)	77 (12%)	314 (48%)
4	393 (8%)	112 (28%)	94 (24%)	187 (48%)
5 or more	518 (10%)	137 (26%)	125 (24%)	256 (49%)





# Emoji as Sentiment Indicators



1. True Sentiment Indication.
2. No Sentiment Indication.
3. Multi-Sentiment Indication.
4. Ambiguous Sentiment Indication.

Figure 1. Examples of the Most Representative Emoji for Each Sentiment in Emoji-Text Dataset.



# 1. True-Sentiment Indication

Sentiment	Tweets
(1) Negative	من تيران وصنافير لسد النهضه يا قلبي احزن علي البلد 🙄 From Tiran and Sanafir islands to Al Nahdha dam, Oh my heart feels sad for the country 🙄
(2) Positive	شفته يا ساره حلوا اوي فعلا يدي تفاؤل وبهجه كده برافو بجد عل ال عمله ده 🥰🥰 Sara, I watched it. It is nice and it really gives the viewers optimism and cheer. Bravo to what he did 🥰🥰
(3) Negative	اطفال سوريا طفوله 🙄 تستنجد انسانيه عالم تواطئ علي دمائمهم 🙄 اغتال احلامهم واغتال ارواحهم 💔 دون ان يرف لهم جفن او يرق لهم قلب Syria's children are the childhood 🙄 that seeks help from the humanity of the world that colludes for their blood 🙄 and assassinates their dreams and souls 💔 without any blink of eye or heartily kindness



## 2. No-Sentiment Indication

(5)  
Positive

امتلات فخرا وانا اقرا النيويورك تايمز وهي تنصح قاده ايران بعدم استفزاز ولي العهد السعودي 🙄 صحيفه لها وزنها العالمي وكتابها عالميين

I was proud when I read the New York Times advises Iran leaders not to provoke the Saudi crown prince 🙄 advice comes from such a well-known newspaper that has a global value and international writers

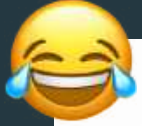
(7)  
Positive

خوي السعد وافي وانا للخوي معكاز وانا محزمه كان الزمن عقد احجاجة نبض العراق 🌱 اخوي الغالي  
عازف ربي يطول عمره 🙏

🙏 ❤️ He is my trustworthy friend and I am his weapon and his wand on which he leans in times of his need. The Iraq's beat 🌱 my beloved brother Aazef, may God prolong his life 🙏



### 3. Multi-Sentiment Indication



(8) Positive نفسي مفاجاه تخليتي منشكح انشكاح منشكجهوش منشكح في تاريخ المنشكحين او اي حاجه تفك عني شويه 😂😂  
I wish I could get a surprise that makes me feel happy in a way that no one felt it in the history of hapeness. Or anything that makes me feel better 😂😂

(11) Negative كريستيانو لاعب مريض نفسيا والله اجل تزعل عشان خويك سجل هدف 😂  
Cristiano, I swear to God, is a psychopathic player, he is upset because his teammate scored a goal 😂



(9) Positive احلا فرحه فرحه النجاح الف مبروك 😎 فرحتي نجاحي  
The best happiness is the happiness of success. Congratulation 😎 My success is my happiness

(12) Negative خليك ببرميل الزباله يا وهيدا بلوك ه 😎  
Keep yourself in the rubbish barrel and here is a block 😎



## 4. Ambiguous Sentiment Indication

(14)  
Negative

بت مش ناقصه رعب انا 🙄 🙄 تثبحي خير وسلميلي اللي وراكي 🤔 🤔 🤔 🤔  
Hay girl, I am already scared 🙄 🙄 Good night and say hi to the one behind you 🤔 🤔  
🤔 🤔



# Work Limitations

- Sentiment behavior of emoji is context-sensitive.
- The source of the text that was analyzed was Twitter only.

# Future Work

- This work will expand upon the analysis presented here by:
  - Comparing the performance of this approach against existing methods.
  - Comparing similar phenomena between Arabic and other different languages and cultures.



Thank You!

