

Laboratorio MAADB

2022-23

Staffieri Michele

Sangalli Attilio

Longo Federica

Indice

- **Intro**
- **Analisi delle risorse**
- **Configurazione**
- **Strutture dati**
- **Popolazione PostgreSQL**
- **Popolazione mongoDB**
- **Analisi & risultati**

Intro: tecnologie utilizzate



Librerie python:

- Pymongo
- Psycopg2
- Wordcloud
- Nltk
- demoji
- Pandas
- matplotlib

Intro: Struttura del progetto

- **ProgettoMAADB**
 - **mongoDB**
 - *mongoConnection.py*
 - *mongoPopulation.py*
 - *mongoAnalysis.py*
 - **postgreSQL**
 - *PGConnection.py*
 - *PGPopulation.py*
 - *PGAnalysis.py*
 - **newResources**
 - ...risorse ed immagini prodotte...
 - **NLPAnalysis**
 - *nlp.py*
 - **utils**
 - ...risorse...

Intro: outline

01 Analisi risorse

02 Configurazione

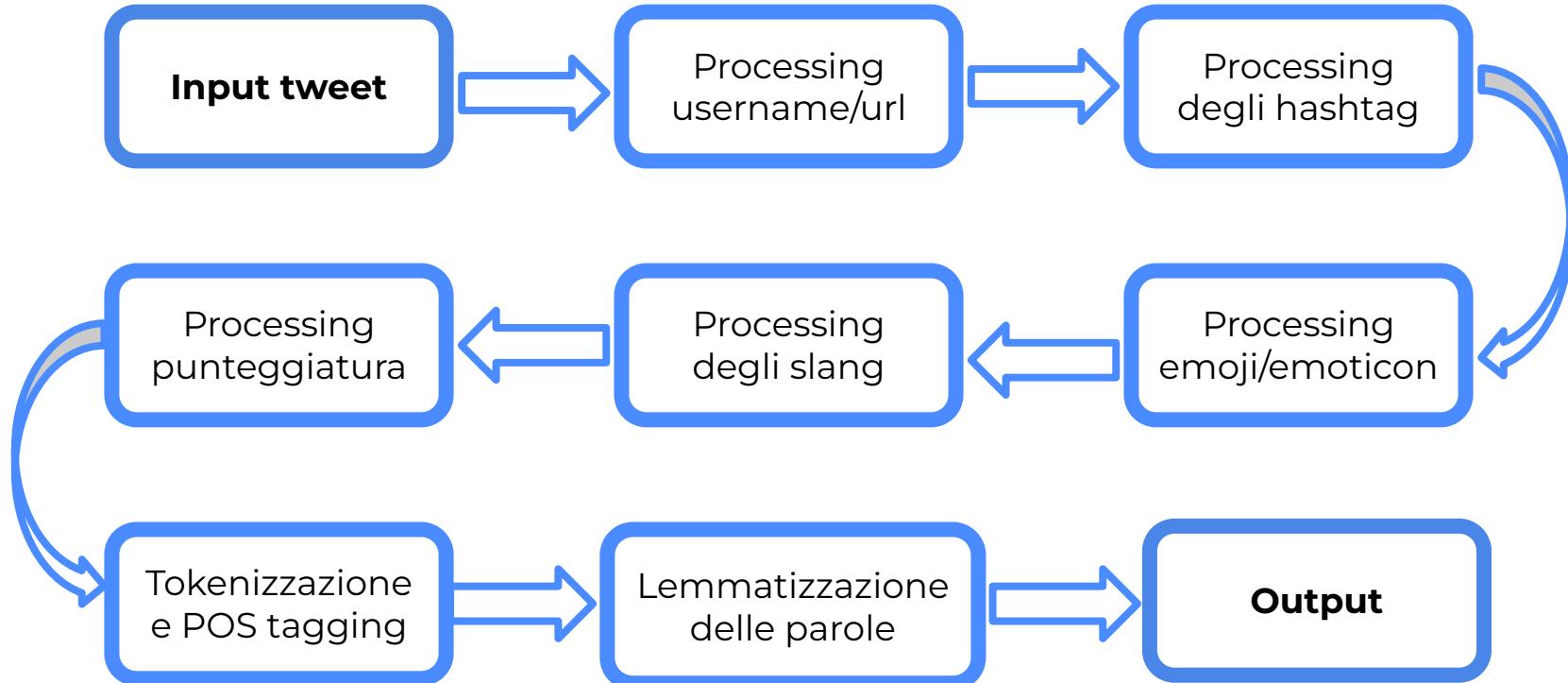
03 Strutture dati

04 Risultati

Analisi delle risorse

- Pipeline NLP per i tweets
- Creazione risorsa lessicale

Processing dei tweets



Processing delle emoticon

```
src > utils > emoji.py > [?] posemoticons
1  posemoticons = ['\:P', '\:p', '\B-', '\:)', '\:-)', '\:)"', '\:"-)', '\:D', '\:-D', '\:\'-)', '\:)"', '\:o)', '\:]', '\:3', '\:c)',
2  '\:)', '\:]', '\=P', '\=p', '\:-P', '\:-p', '\:-b', '\:b', '\:-b', '\:b', '\:-b', '\:b', '\:-b', '\:b', '\:-b', '\:b', '\:-b)', '\:',
3  '\:8)', '\=)', '\:}', '\:^)', '\:8-D', '\:8D', '\:x-D', '\:xD', '\:X-D', '\:xD', '\:-D', '\:D', '\:=-3', '\:=3', '\:B^D', '\:=-)', '\:',
4  '\:^*', '\:^**', '\:( \ ){\' )', '\:^^', '\:(^_^)', '\:^-^', '\:^.^", "\:^B\\\:^", "\:\\L\\\:^", "\:OwO", "\:UwU", "\:O.O", "\:O.O\"", '\:',
5  "\:O.O\\\"", "\:O.O\\\"\\\"", "\:O.O\\\"\\\"\\\"", "\:O.O\\\"\\\"\\\"\\\"", "\:O.O\\\"\\\"\\\"\\\"\\\"", "\:O.O\\\"\\\"\\\"\\\"\\\"\\\"", "\:O.O\\\"\\\"\\\"\\\"\\\"\\\"\\\"", '\:',
6  "\:O.O\\\"\\\"\\\"\\\"\\\"\\\"\\\"", "\:O.O\\\"\\\"\\\"\\\"\\\"\\\"\\\"\\\"", "\:O.O\\\"\\\"\\\"\\\"\\\"\\\"\\\"\\\"", "\:O.O\\\"\\\"\\\"\\\"\\\"\\\"\\\"\\\"", "\:O.O\\\"\\\"\\\"\\\"\\\"\\\"\\\"\\\"", '\:',
7
8  negemoticons = ['\:(', '\:)-(', '\:)'("', "\:)'-", '\:>[', '\:;-c', '\:c', '\:;-<', '\:;<', '\:;-[', '\:][', '\:){', '\:)\'-(', '\:)\\"(',
9  '\:)\\"[', '\:)'(, '\:)[', '\:)'[', '\:)-<', '\:;<', '\:)<', '\:)<', '\:)<', '\:)<', '\:T_T', '\:T.T', '\:(T_T)', '\:y_y', '\:y.y',
10 '\:(Y_Y)', '\:;-;', '\:;-;', '\:;-;', '\:;-;', '\:o ._. o', '\:.-.', '\:;-d', '\:>.<', '\:>.<\\"', '\:>.<\\"', '\:>.<\\"\\\'',
11 '\:>.<\\"', '\:>.<\\"', '\:>.<\\"', '\:>.<\\"', '\:>.<\\"\\\'', '\:>.<\\"\\\'\\\'', '\:>.<\\"\\\'\\\'\\\'', '\:>.<\\"\\\'\\\'\\\'\\\'',
12 '\:.-.', '\:.-\\"', '\:.-.', '\:.-.', '\:;-[', '\:][', '\:;-/', '\:;/', '\:;-\\\'', '\:\\\'', '\:;-()', '\:('', '\:;-c', '\:c', '\:;-<',
13 '\:;<', '\:;-[', '\:][', '\:){', '\:)\'-(', '\:)\\"(', '\:)\\"[', '\:)=(', '\:)[', '\:)=["', '\:)[', '\:)=["', '\:)[', '\:)=["', '\:)[', '\:)=["',
14 '\:)=<', '\:)=<', '\:T_T', '\:T.T', '\:(T_T)', '\:y_y', '\:y.y', '\:(Y_Y)', '\:;-;', '\:;-;', '\:;-;', '\:o ._. o', '\:;-[']
```

Processing degli slang

```
src > utils > 🗄 slang.py > [?] slang_words
1   slang_words = {'afaiak': 'as far as i know', 'afk': 'away from keyboard', 'asap': 'as soon as possible',
2     'atk': 'at the keyboard', 'atm': 'at the moment', 'a3': 'anytime, anywhere, anyplace',
3     'bak': 'back at keyboard', 'bbl': 'be back later', 'bbs': 'be back soon', 'bfn/b4n': 'bye for now',
4     'brb': 'be right back', 'brt': 'be right there', 'btw': 'by the way', 'b4n': 'bye for now',
5     'cu': 'see you', 'cul8r': 'see you later', 'cya': 'see you', 'faq': 'frequently asked questions',
6     'fc': 'fingers crossed', 'fwiw': 'for what it\'s worth', 'fyi': 'for your information',
7     'gal': 'get a life', 'gg': 'good game', 'gmta': 'great minds think alike', 'gr8': 'great!',
8     'g9': 'genius', 'ic': 'i see', 'icq': 'i seek you', 'ilu': 'ilu: i love you',
9     'imho': 'in my honest opinion', 'imo': 'in my opinion', 'iow': 'in other words', 'irl': 'in real life',
10    'kiss': 'keep it simple, stupid', 'ldr': 'long distance relationship', 'lmao': 'laugh my a.. off',
11    'lol': 'laughing out loud', 'ltns': 'long time no see', 'l8r': 'later', 'mte': 'my thoughts exactly',
12    'm8': 'mate', 'nrn': 'no reply necessary', 'oic': 'oh i see', 'pita': 'pain in the a..', 'prt': 'party',
13    'prw': 'parents are watching', 'qpsa?': 'que pasa?', 'rofl': 'rolling on the floor laughing',
14    'roflol': 'rolling on the floor laughing out loud',
15    'rotflmao': 'rolling on the floor laughing my a.. off', 'sk8': 'skate', 'stats': 'your sex and age',
16    'asl': 'age, sex, location', 'thx': 'thank you', 'ttfn': 'ta-ta for now!', 'ttyl': 'talk to you later',
17    'u': 'you', 'u2': 'you too', 'u4e': 'yours for ever', 'wb': 'welcome back', 'wtf': 'what the f...',
18    'wtg': 'way to go!', 'wuf': 'where are you from?', 'w8': 'wait...', '7k': 'sick:-d laugher',
19    'bc': 'because', 'f': 'fuck', '@': 'at', 'tf': 'the fuck', 'af': 'as fuck', 'mf': 'motherfucker',
20    'mfs': 'motherfuckers', 'ASF': 'as fuck', 'bfs': 'best friends', 'oomf': 'one of my friends',
21    'cmon': 'come on', 'fd': 'fucking d..', 'mfer': 'motherfucker', 'iam': 'i am', 'wtfuck': 'what the fuck',
22    'gnite': 'good night', 'fn': 'fake news', 'vs': 'versus', 'kɔŋʒ': 'fuck', 'omg': 'oh my god',
23    'frfr': 'for real for real', 'fr': 'for real', 'im': 'i am', 'idk': 'i don\'t know ', 'idc': 'i don\'t care',
24    'ily': 'i love you'}
```

Esecuzione pipeline NLP

```
'Start analyzing tweets'
'Start analyzing tweet of: Disgust'
End Disgust tweets analysis in: 73.19293022155762 seconds
'Start analyzing tweet of: Anticipation'
End Anticipation tweets analysis in: 70.61226677894592 seconds
'Start analyzing tweet of: Anger'
End Anger tweets analysis in: 72.3595118522644 seconds
'Start analyzing tweet of: Fear'
End Fear tweets analysis in: 68.45676231384277 seconds
'Start analyzing tweet of: Joy'
End Joy tweets analysis in: 69.1069016456604 seconds
'Start analyzing tweet of: Sadness'
End Sadness tweets analysis in: 70.01529955863953 seconds
'Start analyzing tweet of: Surprise'
End Surprise tweets analysis in: 70.59194159507751 seconds
'Start analyzing tweet of: Trust'
End Trust tweets analysis in: 69.43245458602905 seconds
End all tweets analysis in time: 563.7682499885559 seconds
```

Configurazione

- PostgreSQL
- MongoDB

docker-compose.yml

```
1  version: "3"
2
3  services:
4    maadb-ps-db:
5      image: postgres:latest
6      container_name: maadb-ps-db
7      ports:
8        - "5433:5432"
9      env_file:
10        - postgres.env
11      volumes:
12        - database-data:/usr/local/var/postgres
13
14  volumes:
15    database-data:
```

PostgreSQL

Il Database PostgreSQL è stato eseguito in locale utilizzando docker-compose

Versione: latest

Port: 5433

Avvio: docker compose up

postgres.env

| | |
|---|---------------------------------|
| 1 | POSTGRES_HOST_AUTH_METHOD=trust |
| 2 | POSTGRES_USER=maadbsql |
| 3 | POSTGRES_DB=maadbsql |

PostgreSQL

Connessione al db

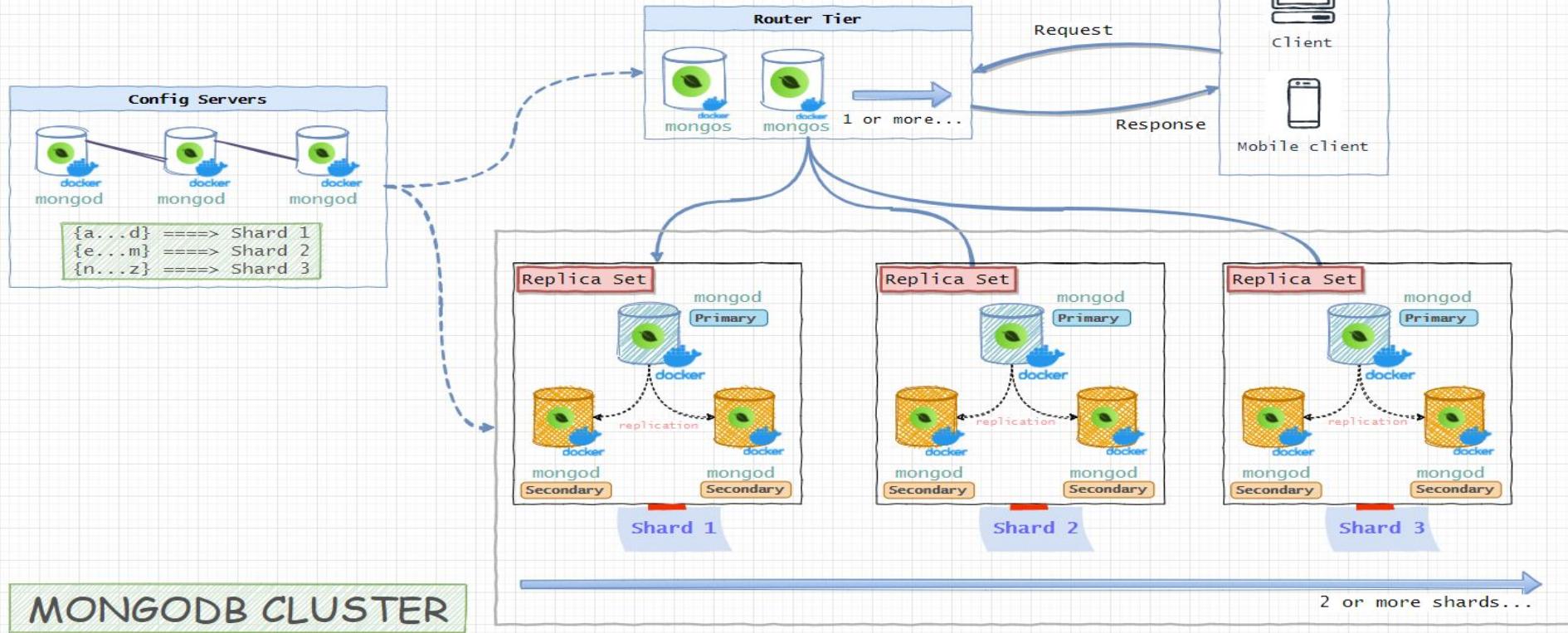
PGConnection.py

```
1 import psycopg2
2
3 4 usages
4
5 class PGConnection:
6
7     def __init__(self):
8         self.conn = None
9         self.create_pg_connection()
10
11     1 usage
12
13     def create_pg_connection(self):
14         try:
15             self.conn = psycopg2.connect('dbname=maadbsql user=maadbsql host=localhost port=5433')
16         except (Exception, psycopg2.DatabaseError) as error:
17             print("Error while postgress connection: ", error)
```

MongoDB

Come per PostgreSQL, eseguiamo il db in locale tramite **docker compose** creando una architettura di questo tipo:

Fonte: <https://github.com/minhhungit/mongodb-cluster-docker-compose>



Mongo router

```
## Router

router01:
  image: mongo:6.0.1
  container_name: router-01
  command: mongos --port 27017 --configdb configsvr01:27017,configsvr02:27017,configsvr03:27017 --bind_ip_all
  ports:
    - 27117:27017
  restart: always
  volumes:
    - ./scripts:/scripts
    - mongodb_cluster_router01_db:/data/db
    - mongodb_cluster_router01_config:/data/configdb

router02:
  image: mongo:6.0.1
  container_name: router-02
  command: mongos --port 27017 --configdb configsvr01:27017,configsvr02:27017,configsvr03:27017 --bind_ip_all
  volumes:
    - ./scripts:/scripts
    - mongodb_cluster_router02_db:/data/db
    - mongodb_cluster_router02_config:/data/configdb
  ports:
    - 27118:27017
  restart: always
  links:
    - router01
```

Mongo config server

```
30 ## Config Servers
31 > configsvr01:
32   image: mongo:6.0.1
33   container_name: mongo-config-01
34   command: mongod --port 27017 --configsvr --replSet rs-config-server
35   volumes:
36     - ./scripts:/scripts
37     - mongodb_cluster_configsvr01_db:/data/db
38     - mongodb_cluster_configsvr01_config:/data/configdb
39   ports:
40     - 27119:27017
41   restart: always
42   links:
43     - shard01-a
44     - shard02-a
45     - shard03-a
```

Configsvr01 < Configsvr02 < Configsvr03

v

Shard01-a
Shard02-a
Shard03-a

Mongo Shard

```
76 ► shard01-a:  
77   image: mongo:6.0.1  
78   container_name: shard-01-node-a  
79   command: mongod --port 27017 --shardsvr --replicaSet rs-shard-01  
80   volumes:  
81     - ./scripts:/scripts  
82     - mongodb_cluster_shard01_a_db:/data/db  
83     - mongodb_cluster_shard01_a_config:/data/configdb  
84   ports:  
85     - 27122:27017  
86   restart: always  
87   links:  
88     - shard01-b  
89     - shard01-c
```



Configurazione sharded collections

Dopo aver avviato i tre componenti della struttura e connessi i vari componenti tra di loro.

```
docker-compose exec router01 mongosh --port 27017
```

Ci connettiamo al primo router

```
use <my_database_name>
```

Crea il nuovo db e ci sposta al suo interno

```
sh.enableSharding("<my_database_name>")
```

Abilita lo sharding sul nuovo db

```
db.createCollection("<collection_name>")
```

Crea una nuova collezione

```
sh.shardCollection("<my_database_name>.<collection_name>",{shard_key: type})
```

Rende la collezione sharded

```
db.<collection_name>.getShardDistribution()
```

Stampa la distribuzione dei dati sugli shards

MongoDB

Connessione al db

MongoConnection.py

```
2 usages
4 class MongoConnection:
5
6     def __init__(self):
7         self.mongo_conn = None
8         self.create_mongo_connection()
9
10    1 usage
11    def create_mongo_connection(self):
12        try:
13            self.mongo_conn = MongoClient(host='mongodb://127.0.0.1:27117,127.0.0.1:27118').progetto
14        except errors.ServerSelectionTimeoutError as err:
15            print("pymongo ERROR:", err)
```

Visualizzazione con mongoDB Compass

Url di accesso: mongodb://127.0.0.1:27117,127.0.0.1:27118/

The screenshot shows the MongoDB Compass interface for a database named 'progettoMaadb'. The left sidebar lists databases: MyDatabase, admin, config, progetto (selected), LexResources3, LexResourcesWords2 (highlighted in green), and Twitter6. The main area displays the 'LexResourcesWords2' collection with 8.3k documents and 3 indexes. The 'Documents' tab is active. A search bar and filter dropdown are present. Below them is an 'Add Data' button and an 'EXPORT DATA' button. The document list shows five entries:

- `_id: ObjectId('64b7e35c355e41f2ea144a61')`
lemma: "accusation"
resources: Array (2)
- `_id: ObjectId('64b7e35c355e41f2ea144a63')`
lemma: "actionable"
resources: Array (2)
- `_id: ObjectId('64b7e35c355e41f2ea144a69')`
lemma: "affront"
resources: Array (5)
- `_id: ObjectId('64b7e35c355e41f2ea144a67')`
lemma: "adder"
resources: Array (4)
- `_id: ObjectId('64b7e35c355e41f2ea144a6b')`
lemma: "alienate"
resources: Array (2)

The 'Find' button is highlighted with a red circle and an arrow pointing to it from the bottom right.

Visualizzazione con mongoDB Compass

Url di accesso: mongodb://127.0.0.1:27117,127.0.0.1:27118/

Explain Plan

Explain provides key execution metrics that help diagnose slow queries and optimize index usage. [Learn more](#)

Visual Tree {} Raw Output

The Explain Plan visualization shows a query flow across three shards:

- Shard Merge:** Returns 8271 documents with an execution time of 4 ms.
- Sharding_Filter:** Returns 2791 documents with 0 ms execution time.
- COLLSCAN:** Returns 2791 documents, examining 2791 documents with 0 ms execution time.
- Sharding_Filter:** Returns 2724 documents with 0 ms execution time.
- COLLSCAN:** Returns 2724 documents, examining 2724 documents with 0 ms execution time.
- Sharding_Filter:** Returns 2756 documents with 0 ms execution time.
- COLLSCAN:** Returns 2756 documents, examining 2756 documents with 2 ms execution time.

Query Performance Summary:

- 8271 documents returned
- 8271 documents examined
- 6 ms execution time
- Is not sorted in memory
- 0 index keys examined
- No index available for this query.

+ - Close

Lo stesso risultato lo si ottiene collegandosi al mongo router e interrogando la collezione

```
docker-compose exec router01 mongosh --port 27017
```

```
[direct: mongos] test> use progetto
switched to db progetto
[direct: mongos] progetto> db.LexResourcesWords2.getShardDistribution()
```

```
{
  data: '1.12MiB',
  docs: 8271,
  chunks: 6,
  'Shard rs-shard-02': [
    '33.11 % data',
    '33.32 % docs in cluster',
    '142B avg obj size on shard'
  ],
  'Shard rs-shard-03': [
    '32.99 % data',
    '32.93 % docs in cluster',
    '143B avg obj size on shard'
  ],
  'Shard rs-shard-01': [
    '33.89 % data',
    '33.74 % docs in cluster',
    '143B avg obj size on shard'
  ]
}
```

In questo caso mostra le percentuali di distribuzione dei dati

Strutture Dati

PostgreSQL

Risorse Lessicali: una tabella per ciascuna emozione: memorizziamo la presenza nei tre tipi di risorse senza score e il loro conteggio totale, e gli score delle risorse con score

| id [PK] | word | w_count | NRC | EmoSN | sentisense | afinn | anewAro | dalActiv |
|---------|---------|---------|-----|-------|------------|-------|---------|----------|
| 1 | hardy | 2 | 1 | 1 | 0 | -1 | 0 | 3 |
| 2 | scream | 1 | 1 | 0 | 0 | 2 | 1 | 1 |
| 3 | darling | 2 | 0 | 1 | 1 | 1 | 2 | 2 |

PostgreSQL

Tweet: una tabella per ogni emozione, memorizziamo la parola, la sua frequenza nell' intero dataset, e la risorsa di riferimento che contiene tale parola.

Inoltre, per ciascuna emozione, anche una tabella per le emoji e una per gli hashtag.

| id [PK] | word | w_count | resources_id [FK] |
|----------------|-------------|----------------|--------------------------|
| 1 | roll | 12 | 3244 |
| 2 | plan | 4 | 1516 |

| id [PK] | emoji | count |
|----------------|-----------------|--------------|
| 1 | :unamused face: | 35 |

Risorse Lessicali: una collezione per memorizzare il nome dei file delle risorse e il numero totale di parole contenute in ciascuno di essi.

Una collezione per memorizzare tutte le parole presenti nelle risorse e la lista di riferimenti alle risorse in cui ogni parola compare

LexResources

```
{  
    _id: resource_filename,  
    sentiment: nome_sentimento,  
    totNumberWords: number_of_words  
}
```

LexResourcesWords

```
{  
    _id: ObjectId("..."),  
    lemma: word,  
    resources : [ {$ref: LexResources, $id: resource_filename}  
                 ,{},...]  
}
```

Tweets: una collezione per memorizzare per ogni emozione, il relativo corpus di tweet. Ogni messaggio di tweet di un certo sentimento è caratterizzato dalla **riga** del file in cui si trova, una lista contenente i **lemmi** del tweet arricchita con **pos, frequenza** e riferimento della stessa parola alla tabella LexResourcesWords. In fine le tre liste di **hashtags, emojis** ed **emoticons** del tweet

Twitter:

```
{  
  sentiment : "joy"|"sadness"|"trust"|"anger"|"anticipation"|"disgust"|"fear"|"surprise"},  
  {doc_number : order_number_of_the_document_in_the_sentiment_file,  
   words : [{lemma : "the_lemma", POS : posTag, freq : count_occurrences_in_msg,  
             in_lex_resources : {$ref : LexResourcesWords, $id : reference_Object_id}  
             },  
           ...]  
   hashtags : [list_of_hashtags],  
   emojis : [list_of_emojis],  
   emoticons : [list_of_emoticons]},  
   ...  
 }
```



Popolazione del database PosgreSQL

Scelte

Per le tabelle dei tweets (come richiesto dalla traccia) è stato imposto un vincolo di chiave esterna dalla word della riga all'id della stessa parola nella tabella delle risorse lessicali, null se non c'è

Esecuzione e tempi



Creazione tabelle 8 delle **risorse lessicali** in 1.59 secondi

Creazione delle 8 tabelle dei **tweets** in 61.12 secondi

Creazione delle 8 tabelle **hashtags** in 10.96 secondi

Creazione delle 8 tabelle **emojis** in 0.63 secondi

Creazione delle 8 tabelle **emoticons** in 0.08 secondi

```
'You choose PostgreSQL'  
Resources created in: 1.5940003395080566  
Twitter created in: 61.12205386161804  
Hashtag created in: 10.962129592895508  
Emoji created in: 0.6378324031829834  
Emoticons created in: 0.08719968795776367
```

Popolazione del database MongoDB

Scelte - Tweets in chunks



```
_id: ObjectId('64bbded77d3e71b1d635f15a')
sentiment: "Anticipation"
▶ content: Array (20000)
```

```
_id: ObjectId('64bbded77d3e71b1d635f15b')
sentiment: "Anticipation"
▶ content: Array (20000)
```

```
_id: ObjectId('64bbded77d3e71b1d635f15c')
sentiment: "Anticipation"
▶ content: Array (20000)
```

Come da documentazione, (<https://www.mongodb.com/docs/manual/reference/limits/>) un documento in MongoDB può essere di al massimo **16MB**

Inserendo in un singolo documento tutti i dati dei 60k tweets si andava a superare questa soglia.

Per risolvere questo problema i tweets di ogni sentimento sono divisi in tre chunks da **20k** tweets l'uno facilmente identificabili e raggruppabili tramite il loro campo **'sentiment'**.

Si va cioè a spezzare in tre parti la struttura proposta precedentemente ([qui](#))

Scelte - Shard keys

Per la collezione **LexResources** è stato scelto il campo
‘_id’ di tipo 1

Per la collezione **LexResourcesWords** è stato scelto il
campo ‘lemma’ di tipo 1

Per la collezione **Twitter** è stato scelto il campo
‘sentiment’ di tipo hashed

Esecuzione e tempi

Creazione collezione **LexResources** in: 0.04 secondi

Creazione collezione **LexResourcesWords** in 1.27 secondi

Creazione della collezione **Twitter** in 20.10 secondi

```
'You choose mongoDB'  
LexResources created in: 0.04031205177307129  
LexResourcesWords created in: 1.2711999416351318  
Twitter created in: 22.108475923538208
```

Analisi & Risultati

Cosa è stato prodotto

Analisi in PostgreSQL

- Istogrammi
- Lista delle nuove parole
- Nuova risorsa lessicale (risorse lessicali + nuove parole)
- Statistiche sulla presenza delle parole tra tweets e risorse
- Wordclouds

Analisi MongoDB

- Wordclouds
- Statistiche sulla presenza delle parole tra tweets e risorse
- Top 10 token (parole, emoji, hashtag, emoticon)

(* con poche modifiche all' aggregation pipeline per le statistiche è possibile ottenere anche le nuove parole)

Analisi MongoDB - frequenze

Analisi delle frequenze di hashtag, emoji ed emoticons, per tutti i sentimenti

```
'Starting hashtags analysis'  
hashtags calculated in: 7.201111793518066  
'Starting emojis analysis'  
emojis calculated in: 111.03979659080505  
'Starting emoticons analysis'  
emoticons calculated in: 4.296820402145386
```

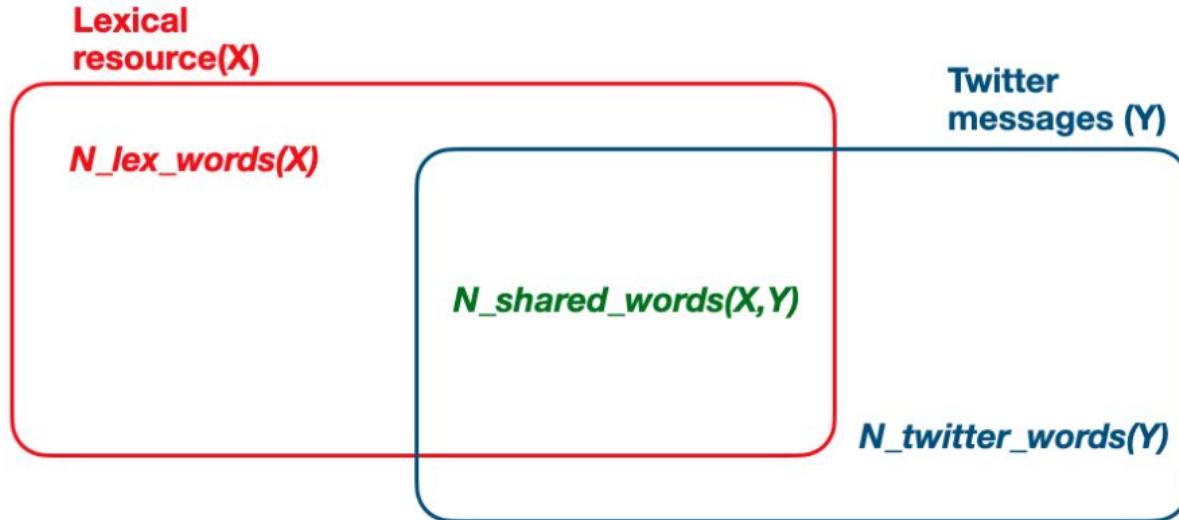
Analisi delle frequenze delle parole, per ciascun sentimento

```
'Starting words frequency analysis'  
Disgust calculated in: 99.6001386642456
```

Tendenzialmente fra i 96 e i 110 secondi

Analisi MongoDB - intersezioni

Viene calcolata l'intersezione fra le parole dei tweet e quelle di uno specifico file di risorsa lessicale, per uno stesso sentimento



$$perc_presence_lex_res (X, Y) = \frac{N_shared_words(X, Y)}{N_lex_words(X)}$$

$$perc_presence_twitter (X, Y) = \frac{N_shared_words(X, Y)}{N_twitter_words(Y)}$$

Analisi MongoDB - intersezioni

Tempi di esecuzione del calcolo delle intersezioni

```
'Starting words in tweets and resources analysis'  
Statistics for Disgust calculated in: 111.81236815452576  
Statistics for Anticipation calculated in: 111.09495878219604  
Statistics for Anger calculated in: 119.82070970535278  
Statistics for Fear calculated in: 108.2624146938324  
Statistics for Joy calculated in: 111.29823279380798  
Statistics for Sadness calculated in: 111.56412839889526  
Statistics for Surprise calculated in: 116.03676795959473  
Statistics for Trust calculated in: 104.33899259567261
```

Parole delle risorse lessicali nei tweet

| | | | |
|---------------------------|--------|-----------------------|--------|
| Anger - EmoSN | 58.19% | Joy - EmoSN | 62.66% |
| Anger - NRC | 46.99% | Joy - NRC | 59.80% |
| Anger - sentisense | 56.36% | Joy - sentisense | 45.31% |
| Anticipation - NRC | 50.42% | Sadness - NRC | 42.15% |
| Anticipation - sentisense | 70.14% | Sadness - sentisense | 40.91% |
| Disgust - NRC | 45.18% | Surprise - NRC | 56.37% |
| Disgust - sentisense | 51.87% | Surprise - sentisense | 55.17% |
| Fear - NRC | 40.51% | Trust - NRC | 49.72% |
| Fear - sentisense | 61.29% | | |

Parole dei tweet nelle risorse lessicali

| | | | |
|---------------------------|--------|-----------------------|--------|
| Anger - EmoSN | 0.681% | Joy - EmoSN | 6.898% |
| Anger - NRC | 1.938% | Joy - NRC | 1.215% |
| Anger - sentisense | 0.102% | Joy - sentisense | 0.171% |
| Anticipation - NRC | 1.459% | Sadness - NRC | 1.816% |
| Anticipation - sentisense | 0.348% | Sadness - sentisense | 0.195% |
| Disgust - NRC | 1.584% | Surprise - NRC | 0.856% |
| Disgust - sentisense | 0.921% | Surprise - sentisense | 0.045% |
| Fear - NRC | 2.173% | Trust - NRC | 1.865% |
| Fear - sentisense | 0.345% | | |

Analisi PostgreSQL

I dati vengono scaricati dal database PostgreSQL tramite query e analizzati:

- Per ogni sentimento vengono generate due nuove risorse:
 - Una contenente tutte le nuove parole dei tweets e le loro frequenze
 - Una contenente solo quelle nuove rispetto alle risorse
- Come per l'analisi in MongoDB sono stati calcolati gli insiemi intersezione delle parole presenti sia nei tweets che nelle risorse lessicali. Anche qui sono state calcolate le statistiche e salvate in appositi file
- L'analisi ha portato anche alla generazione di histogrammi che rappresentano le percentuali di parole dei tweet di un certo sentimento presenti nelle risorse lessicali di altri sentimenti
- Tramite le frequenze memorizzate nelle tabelle di PostgreSQL sono state generate delle wordclouds

Nelle prossime slides sono riportati:

- Le top 10 parole, emoji, hashtag ed emoticons prodotte dall'analisi in mongoDB
- Le WordCloud prodotte dall'analisi di mongoDB
- Gli istogrammi prodotti dall'analisi in PostgreSQL

Nel progetto è possibile ispezionare risorse lessicali prodotte e le wordclouds prodotte con PostgreSQL

Top 10 - Anger

| word | freq |
|--------|------|
| get | 7063 |
| go | 5553 |
| fuck | 3215 |
| know | 2999 |
| hate | 2522 |
| want | 2514 |
| laugh | 2424 |
| people | 2359 |
| loud | 2354 |
| make | 2335 |

| hashtag | freq |
|------------|------|
| ugh | 1698 |
| annoying | 1218 |
| annoyed | 1173 |
| fuckyou | 1171 |
| pissed | 1087 |
| ughh | 401 |
| frustrated | 307 |
| p | 303 |
| job | 283 |
| jobs | 218 |

Top 10 - Anger

| emoji | freq |
|---------------------------|-------|
| unamused face | 23514 |
| pouting face | 16550 |
| angry face | 5163 |
| confounded face | 3842 |
| pensive face | 1487 |
| face with steam from nose | 1157 |
| thumbs down | 1029 |
| loudly crying face | 977 |
| oncoming fist | 964 |
| water pistol | 933 |

| emoticon | freq |
|----------|------|
| >.< | 3992 |
| >_< | 1196 |
| :) | 352 |
| :3 | 254 |
| :(| 243 |
| -.- | 149 |
| -_- | 125 |
| :-) | 96 |
| ;) | 60 |
| :c | 45 |

Top 10 - Anticipation

| word | freq |
|-------|------|
| get | 7764 |
| go | 7058 |
| know | 3716 |
| day | 3306 |
| see | 2967 |
| laugh | 2946 |
| loud | 2871 |
| want | 2773 |
| work | 2730 |
| feel | 2622 |

| hashtag | freq |
|-----------------|------|
| excited | 3732 |
| cantwait | 2916 |
| soexcited | 1545 |
| pumped | 927 |
| waiting | 358 |
| tooexcited | 280 |
| oomf | 255 |
| sopumped | 170 |
| confessionnight | 150 |
| exciting | 131 |

Top 10 - Anticipation

| emoji | freq |
|--------------------------------|------|
| confused face | 8138 |
| beaming face with smiling eyes | 587 |
| red heart | 462 |
| pensive face | 453 |
| smiling face with smiling eyes | 394 |
| smiling face with heart eyes | 382 |
| thumbs up | 318 |
| thumbs down | 309 |
| neutral face | 301 |
| unamused face | 261 |

| emoticon | freq |
|-----------|-------|
| :/ | 37197 |
| :-/ | 1990 |
| :) | 1753 |
| :\ :\\ | 1158 |
| :(:(| 720 |
| :-\ | 559 |
| :3 | 258 |
| ;) | 167 |
| :p | 162 |
| -- | 123 |

Top 10 - Disgust

| word | freq |
|-------|------|
| get | 6873 |
| go | 5333 |
| know | 3364 |
| laugh | 2992 |
| loud | 2781 |
| want | 2480 |
| make | 2272 |
| think | 2179 |
| say | 2154 |
| fuck | 2120 |

| hashtag | freq |
|---------|------|
| bored | 1066 |
| gross | 702 |
| ew | 556 |
| rude | 481 |
| job | 356 |
| eww | 326 |
| nasty | 305 |
| jobs | 272 |
| yuck | 262 |
| oomf | 258 |

Top 10 - Disgust

| emoji | freq |
|------------------------|-------|
| unamused face | 19289 |
| smirking face | 12161 |
| expressionless face | 6177 |
| confounded face | 3355 |
| persevering face | 3229 |
| neutral face | 2862 |
| face with tears of joy | 1794 |
| pensive face | 1498 |
| water pistol | 836 |
| loudly crying face | 790 |

| emoticon | freq |
|----------|------|
| >.< | 3286 |
| >_< | 1010 |
| :) | 352 |
| :3 | 255 |
| :(" | 182 |
| --. | 101 |
| -_- | 92 |
| ;) | 83 |
| :-) | 69 |
| :c | 41 |

Top 10 - Fear

| word | freq |
|-------|------|
| get | 6304 |
| go | 5015 |
| know | 3436 |
| laugh | 2984 |
| love | 2806 |
| loud | 2793 |
| see | 2591 |
| day | 2581 |
| make | 2538 |
| good | 2445 |

| hashtag | freq |
|-------------------|------|
| awkward | 1154 |
| awk | 422 |
| decisions | 387 |
| thatawkwardmoment | 358 |
| scared | 358 |
| scary | 309 |
| stressed | 274 |
| creepy | 248 |
| in6thgrade | 240 |
| nervous | 220 |

Top 10 - Fear

| emoji | freq |
|--------------------------------|-------|
| smiling face with smiling eyes | 22072 |
| flushed face | 14620 |
| face screaming in fear | 6288 |
| red heart | 2860 |
| face with tears of joy | 2592 |
| smiling face with heart eyes | 2300 |
| face blowing a kiss | 2146 |
| beaming face with smiling eyes | 2093 |
| thumbs up | 1944 |
| anxious face with sweat | 1490 |

| emoticon | freq |
|-----------|-------|
| :/ | 13804 |
| :-/ | 741 |
| :) | 681 |
| :\ :\\ | 420 |
| :(:(\ | 327 |
| :-\ | 192 |
| :3 | 169 |
| ;) | 82 |
| :p | 68 |
| -.- | 64 |

Top 10 - Joy

| word | freq |
|-------|------|
| get | 5868 |
| love | 5502 |
| laugh | 4613 |
| go | 4207 |
| loud | 4063 |
| good | 3477 |
| know | 2874 |
| make | 2620 |
| see | 2470 |
| day | 2469 |

| hashtag | freq |
|-----------------|------|
| love | 833 |
| confessionnight | 529 |
| winning | 403 |
| happy | 277 |
| oomf | 165 |
| yay | 151 |
| orlando | 116 |
| happyweet | 92 |
| loveyou | 80 |
| family | 79 |

Top 10 - Joy

| emoji | freq |
|---------------------------------|-------|
| face with tears of joy | 14013 |
| smiling face with heart-eyes | 6493 |
| smiling face | 2611 |
| face blowing a kiss | 2017 |
| red heart | 1950 |
| winking face with tongue | 1428 |
| grinning face with smiling eyes | 1345 |
| grinning face with big eyes | 1192 |
| squinting face with tongue | 1086 |
| face savoring food | 1026 |

| emoticon | freq |
|----------|-------|
| :) | 21468 |
| :-) | 3793 |
| :3 | 2906 |
| :p | 2243 |
| ^.^ | 596 |
| =) | 490 |
| :') | 305 |
| ;) | 298 |
| :b | 228 |
| :/ | 155 |

Top 10 - Sadness

| word | freq |
|------|------|
| go | 6455 |
| get | 6337 |
| miss | 4665 |
| want | 3277 |
| know | 3273 |
| see | 2874 |
| wish | 2812 |
| feel | 2516 |
| work | 2490 |
| day | 2469 |

| hashtag | freq |
|-----------------|------|
| fml | 1877 |
| sadtweet | 1022 |
| confessionnight | 366 |
| oomf | 165 |
| in6thgrade | 163 |
| depressed | 141 |
| sadday | 137 |
| depressing | 126 |
| sad | 106 |
| fuckmylife | 80 |

Top 10 - Sadness

| emoji | freq |
|-----------------------|-------|
| pensive face | 10741 |
| loudly crying face | 9895 |
| crying face | 3757 |
| disappointed face | 3532 |
| broken heart | 3134 |
| persevering face | 1973 |
| sad but relieved face | 1751 |
| unamused face | 1058 |
| weary face | 1018 |
| tired face | 693 |

| emoticon | freq |
|----------|-------|
| :(" | 24093 |
| :'(| 2209 |
| :-(" | 1273 |
| :c | 1026 |
| :/ | 338 |
| :[| 128 |
| -.- | 105 |
| ;) | 91 |
| -_- | 76 |
| :{ | 63 |

Top 10 - Surprise

| word | freq |
|-------|------|
| get | 5982 |
| go | 4594 |
| laugh | 4561 |
| loud | 4168 |
| know | 3263 |
| say | 2739 |
| think | 2635 |
| f | 2231 |
| see | 2148 |
| god | 2120 |

| hashtag | freq |
|-------------|------|
| wtf | 8687 |
| weird | 2963 |
| wow | 2582 |
| crazy | 2430 |
| omg | 2164 |
| random | 1794 |
| oomf | 247 |
| blackfriday | 198 |
| lol | 184 |
| love | 119 |

Top 10 - Surprise

| emoji | freq |
|--------------------------------|------|
| astonished face | 4539 |
| knocked-out face | 4177 |
| face with open mouth | 3056 |
| hushed face | 2281 |
| water pistol | 1287 |
| face with tears of joy | 1181 |
| flushed face | 1124 |
| face screaming in fear | 899 |
| smiling face with heart-eyes | 567 |
| beaming face with smiling eyes | 501 |

| emoticon | freq |
|----------|------|
| :) | 698 |
| :p | 269 |
| :(" | 257 |
| :3 | 242 |
| ;) | 197 |
| -.- | 192 |
| -_- | 140 |
| >.< | 126 |
| ^.^ | 88 |
| :c | 76 |

Top 10 - Trust

| word | freq |
|-------|-------|
| love | 10473 |
| get | 4924 |
| go | 3547 |
| good | 2860 |
| make | 2792 |
| know | 2726 |
| day | 2591 |
| see | 2511 |
| laugh | 2480 |
| night | 2283 |

| hashtag | freq |
|-----------------|------|
| love | 1520 |
| confessionnight | 430 |
| bestfriend | 342 |
| oomf | 201 |
| loveyou | 177 |
| sexy | 163 |
| family | 143 |
| loveher | 116 |
| cute | 110 |
| happy | 100 |

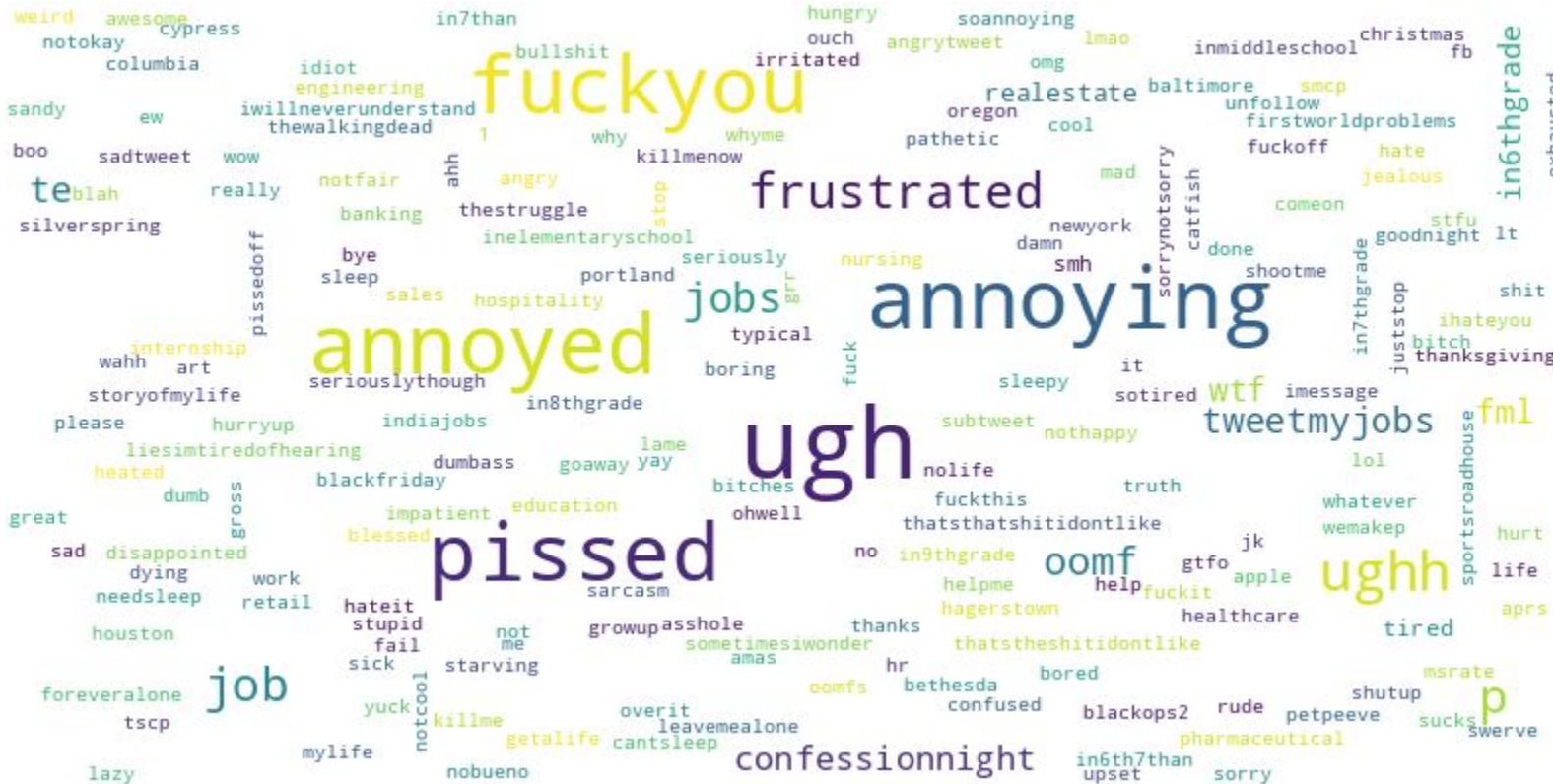
Top 10 - Trust

| emoji | freq |
|--------------------------------|-------|
| red heart | 17401 |
| smiling face with heart-eyes | 11200 |
| face blowing a kiss | 9488 |
| heart suit | 4695 |
| purple heart | 4136 |
| blue heart | 4055 |
| winking face | 3489 |
| yellow heart | 2588 |
| winking face with tongue | 2555 |
| smiling face with smiling eyes | 2267 |

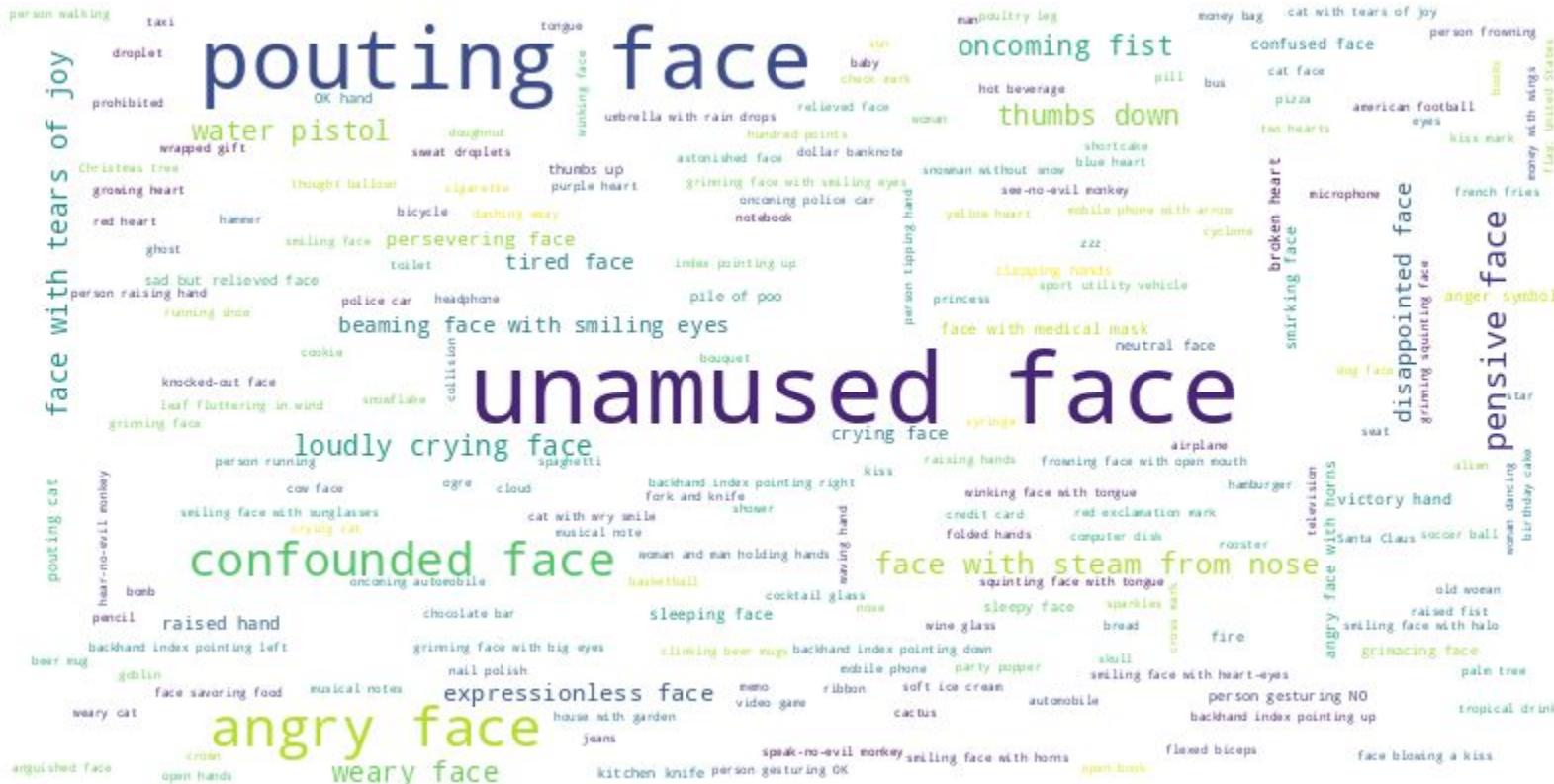
| emoticon | freq |
|----------|------|
| :p | 4027 |
| :) | 2190 |
| :b | 384 |
| ;) | 350 |
| :-p | 271 |
| :(" | 210 |
| :-) | 179 |
| :3 | 152 |
| =p | 131 |
| :_ | 119 |

Wordcloud - Anger (word)

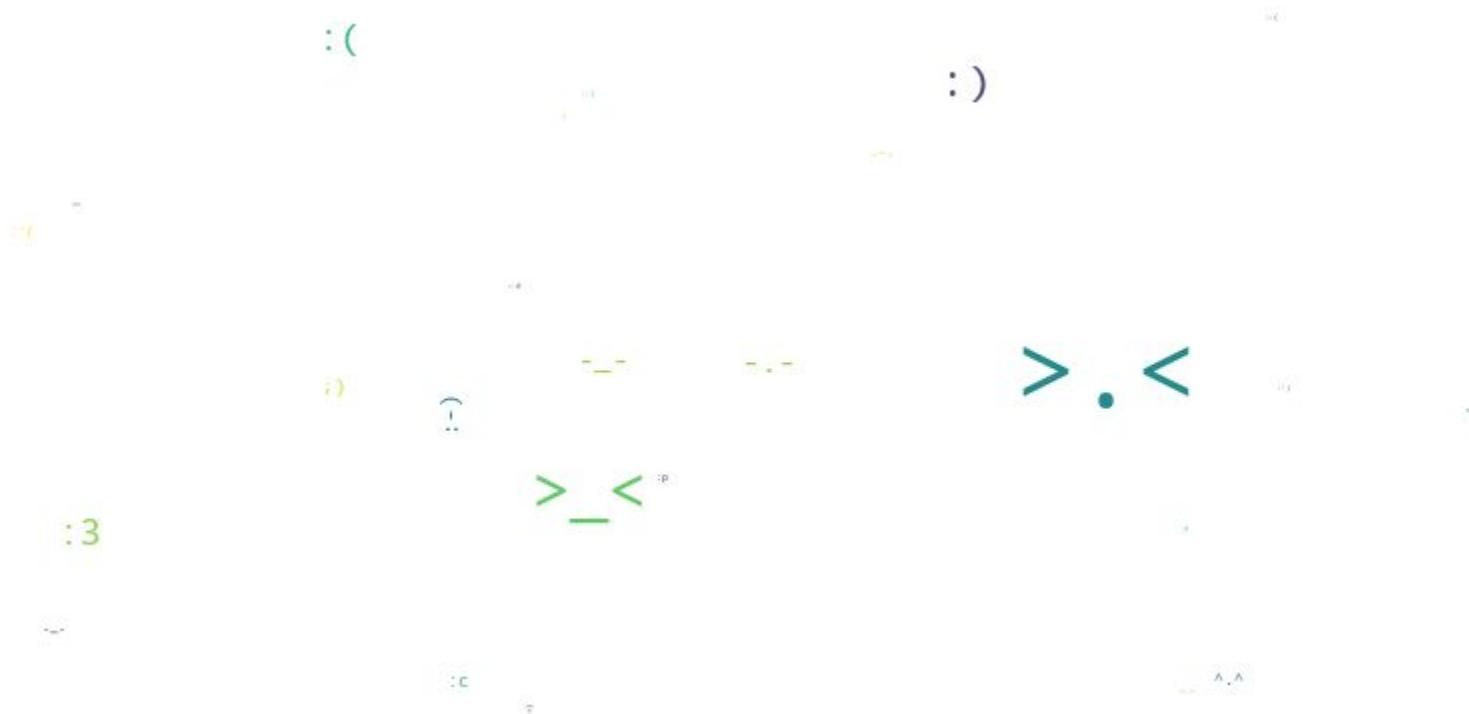
Wordcloud - Anger (hashtag)



Wordcloud - Anger (emoji)



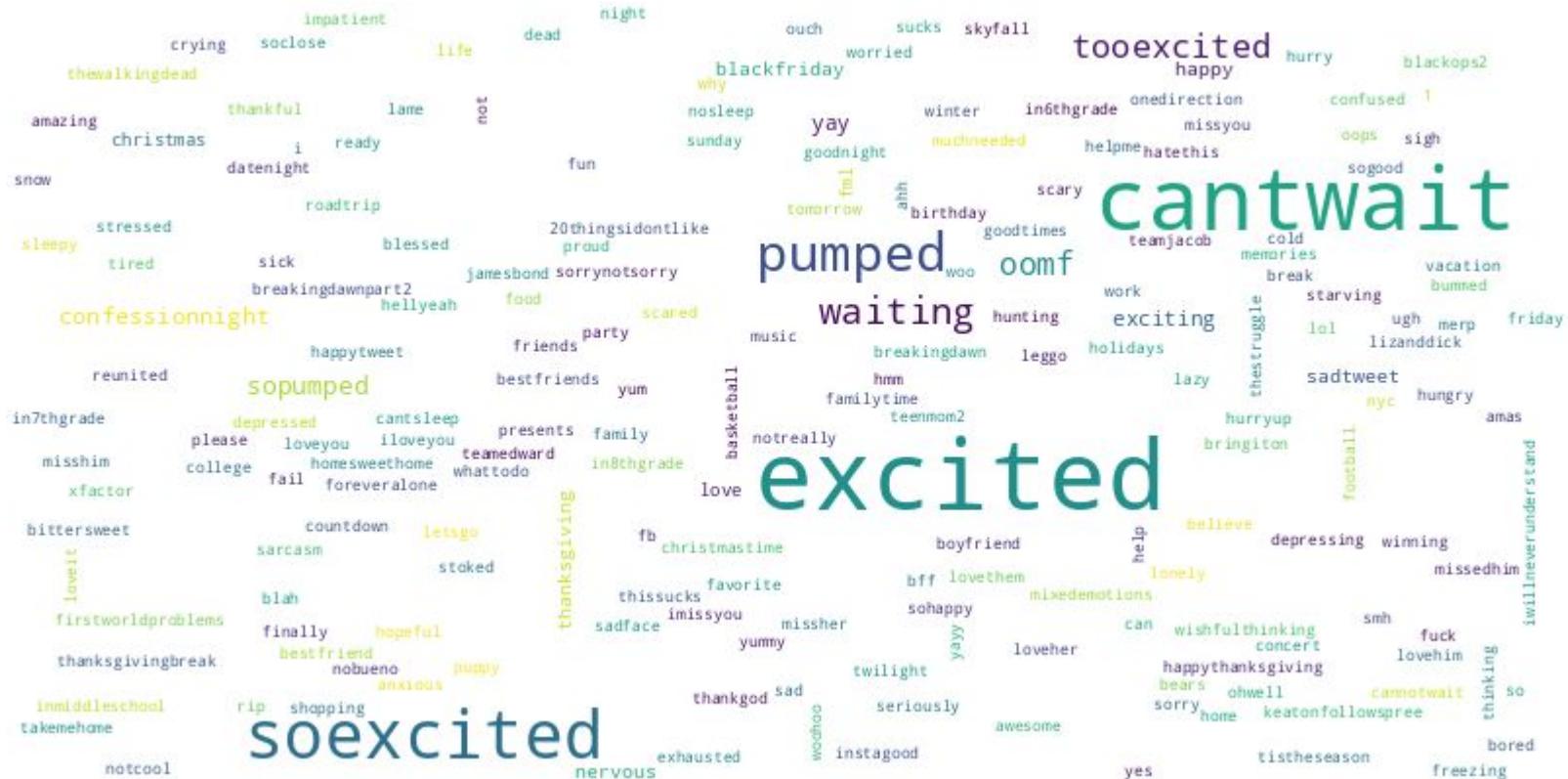
Wordcloud - Anger (emoticon)



Wordcloud - Anticipation (word)

A wordcloud visualization showing the frequency of words related to anticipation. The words are in various colors and sizes, with larger words appearing more frequently. Key words include 'get', 'work', 'day', 'good', 'loud', 'tomorrow', 'time', 'back', 'text', 'find', 'well', 'know', 'say', 'oh', 'home', 'life', 'way', 'text', 'find', 'well', 'ill', 'hour', 'morning', 'cantwait', 'friday', 'early', 'take', 'meet', 'long', 'yet', 'bore', 'nothing', 'mean', 'look', 'feeling', 'hate', 'new', 'look', 'help', 'class', 'hurt', 'keep', 'never', 'hard', 'eat', 'suck', 'cant', 'make', 'great', 'late', 'cry', 'yeah', 'wait', 'bed', 'give', 'day', 'ni', 'thing', 'maybe', 'get', 'head', 'first', 'think', 'still', 'live', 'sadguy', 'someone', 'pretty', 'wrong', 'god', 'christmas', 'month', 'want', 'cause', 'lotgotta', 'try', 'family', 'break', 'soon', 'even', 'almost', 'everyone', 'haha', 'next', 'feel', 'game', 'week', 'good', 'already', 'work', 'start', 'real', 'bad', 'lose', 'Leave', 'finally', 'through', 'tho', 'school', 'part', 'fall', 'buy', 'aww', 'ask', 'sick', 'drive', 'stuff', 'start', 'real', 'bad', 'lose', 'Leave', 'finally', 'through', 'tho', 'school', 'part', 'let', 'phone', 'sister', 'actually', 'use', 'asleep', 'loud', 'need', 'thanksgiving', 'wanna', 'ready', 'baby', 'laugh', 'cold', 'so', 'tomorrow', 'house', 'talk', 'old', 'girl', 'miss', 'say', 'oh', 'home', 'mom', 'time', 'follow', 'anything', 'really', 'damn', 'still', 'O', 'ever', 'last', 'please', 'show', 'know', 'life', 'big', 'back', 'tell', 'stop', 'best', 'do', 'bo', 'call', 'come', 'see', 'fuck', 'tire', 'alone', 'away', 'way', 'dont', 'as', 'much', 'tonight', 'tshit', 'today', 'play', 'text', 'find', 'well', 'person', 'black', 'weekend', 'tweet', 'stay', 'smh', 'right', 'sure', 'man', 'birthday', 'excite', 'food', 'love', 'hope', 'something', 'movie'

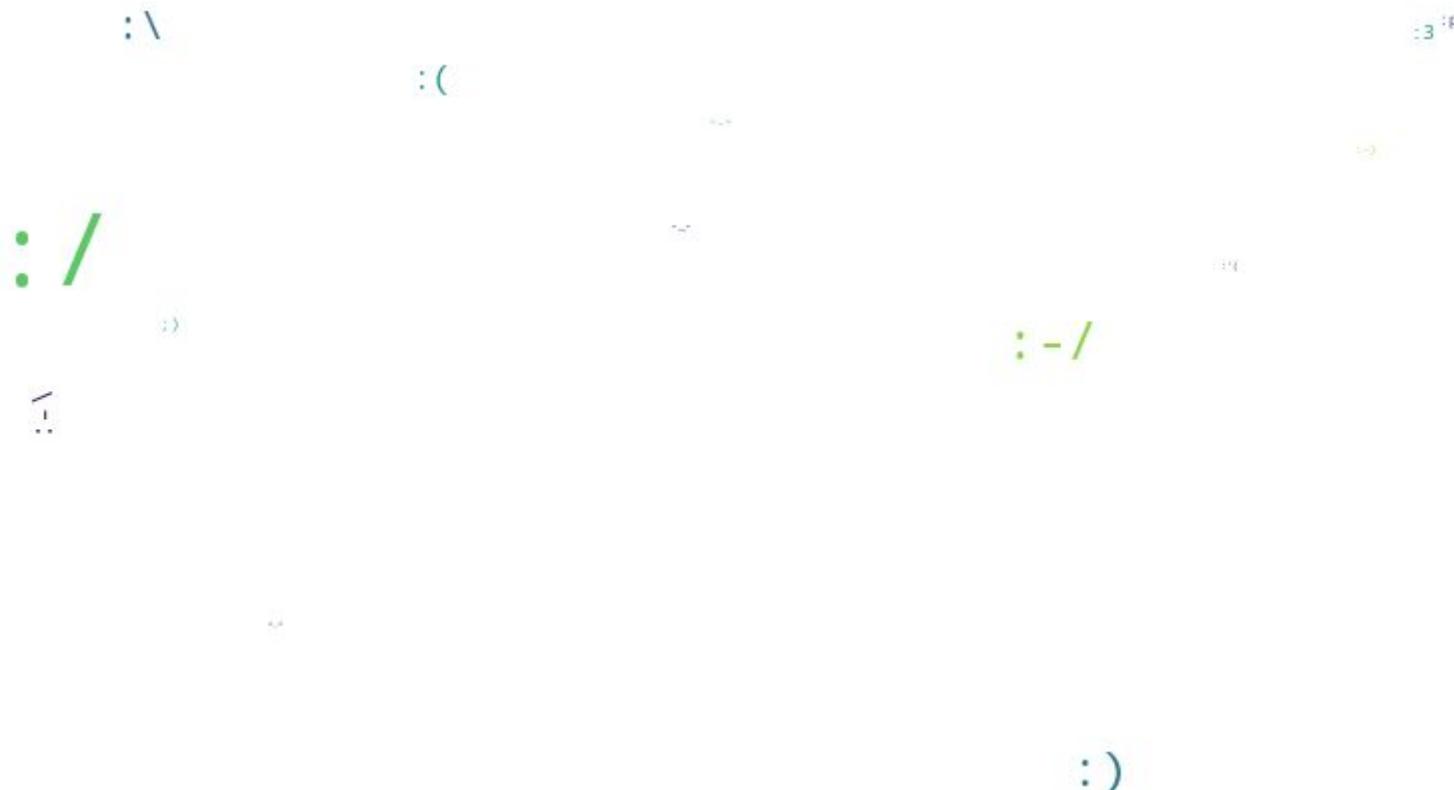
Wordcloud - Anticipation (hashtag)



Wordcloud - Anticipation (emoji)



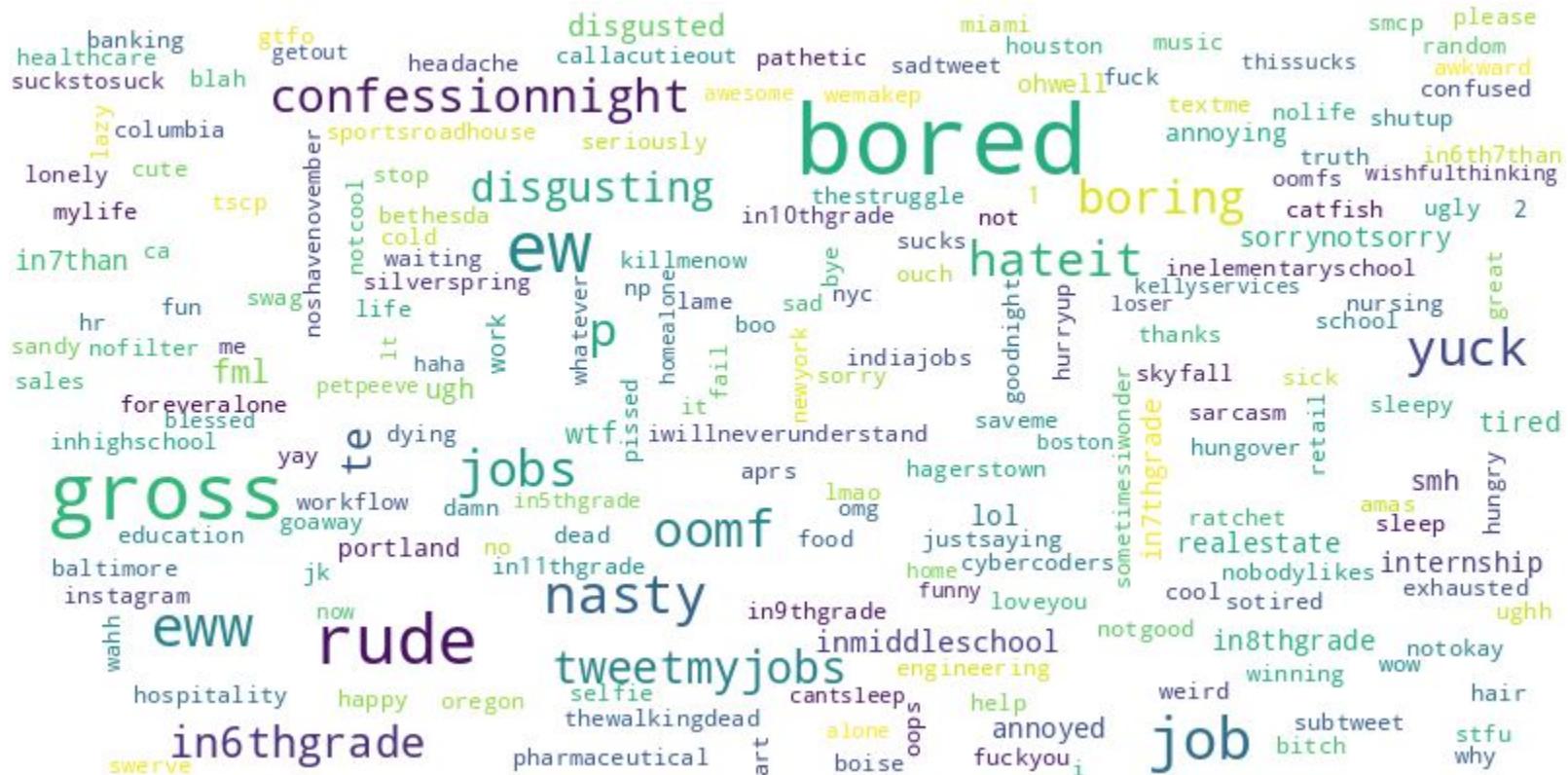
Wordcloud - Anticipation (emoticon)



Wordcloud - Disgust (word)

guess ready twitter life break
say gottaday hurt next best fuck tho
find way week start tomorrow wear cute as
dx walk many start tomorrow
hit see baby look stay want shit stop p
see mom first follow wanna much love kill
okay yet bitch class good dont happy keep sick still
mom suck hell give boy today car phone face
yeah last please tell name picture take someone always
last get care try already home everyone help think leave
try show school dad hard smh wait night
brother mad long sit sorry room
play black bed tweet sleep stupid never hair go tire drive
talk happen bore morning big wake
text turn people gonna real ever well need cause person
loud away old make movie head fall do even guy
year turn wish cool new live boyfriend
missuse god rt come sister time girl damn game
tonight use bad hour something end feel back job know
use everything weekend seriously actually oh nothing watch

Wordcloud - Disgust (hashtag)



Wordcloud - Disgust (emoji)



Wordcloud - Disgust (emoticon)

> . <

;

:)

> _ <

:)

: (

:3

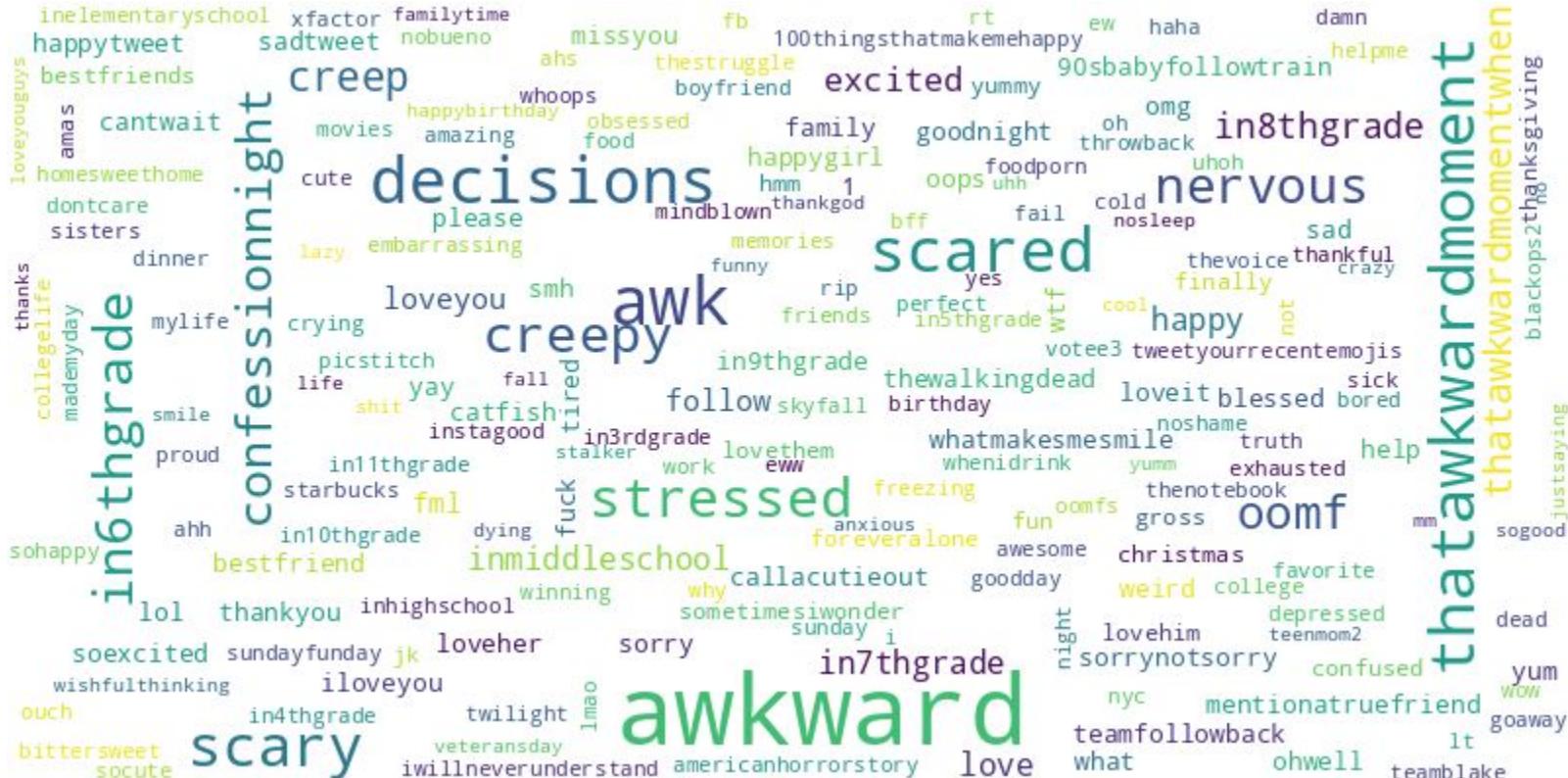
Wordcloud - Fear (word)

A wordcloud visualization where the size of each word represents its frequency or importance. The words are colored in various shades of green, blue, and yellow. Some words have small descriptive suffixes in smaller text below them.

Key words include:

- hour, sick, really, look, soon, miss, yeah, god, little, work, people, happy, pretty, come, day, want, way, much, room, mom, suck, well, see, keep, great, turn, still, stay, make, watch, hard, ill, try, cause, shit, find, school, already, happen, even, actually, start, thanksgiving, friday, fuck, moment, show, loud, weekend, let, sure, night, buy, leave, sister, never, long, need, lose, wish, cute, leave, old, hurt, call, best, right, person, person, mean, hair, play, world, dinner, tomorrow, thanks, oh, girl, please, go, sit, give, face, baby, bad, year, bitch, crazy, week, talk, life, birthday, get, know, think, take, almost, seriously, walk, run, damn, okay, man, thank, hahaha, sorry, haha, car, time, gonna, as, last, remember, christmas, everything, home, laugh, glad, tell, next, boyfriend, wait, back.

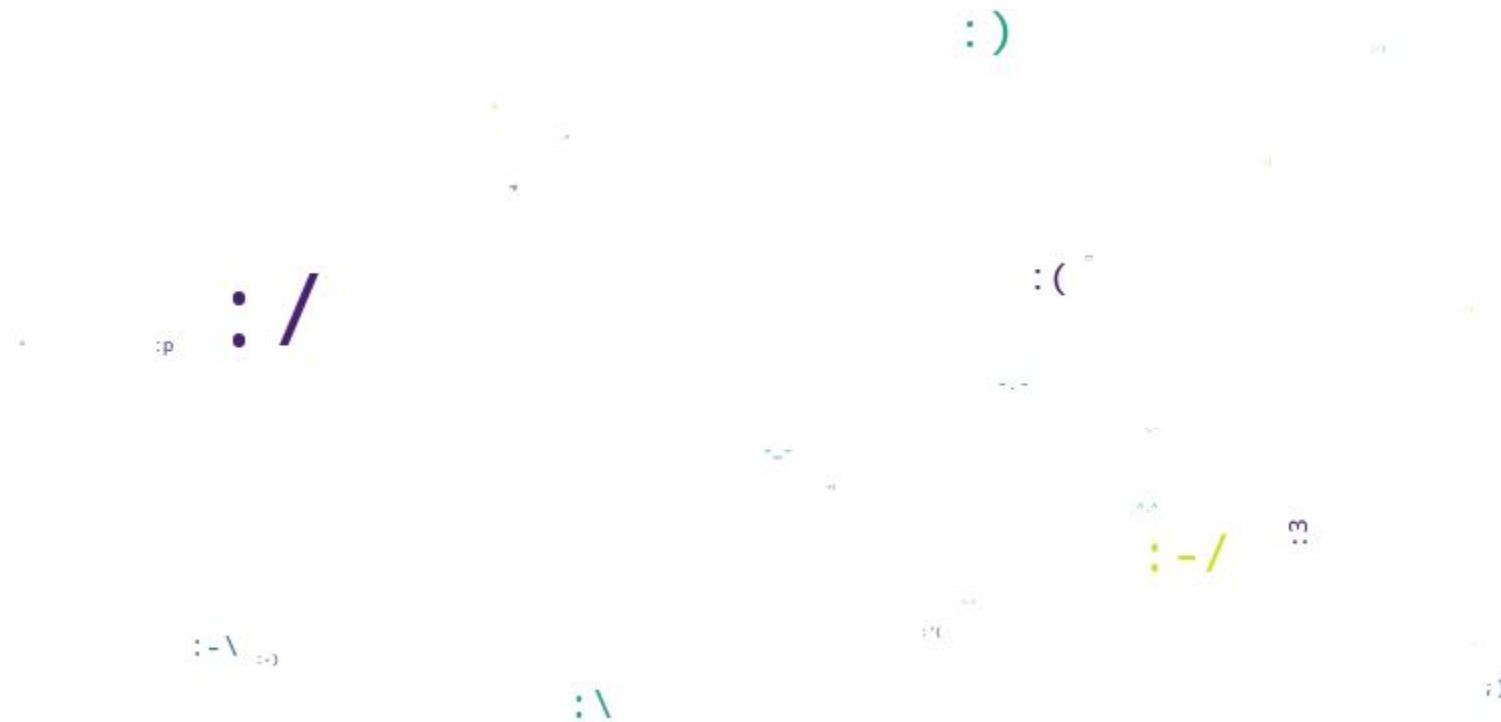
Wordcloud - Fear (hashtag)



Wordcloud - Fear (emoji)



Wordcloud - Fear (emoticon)



Wordcloud - Joy (word)

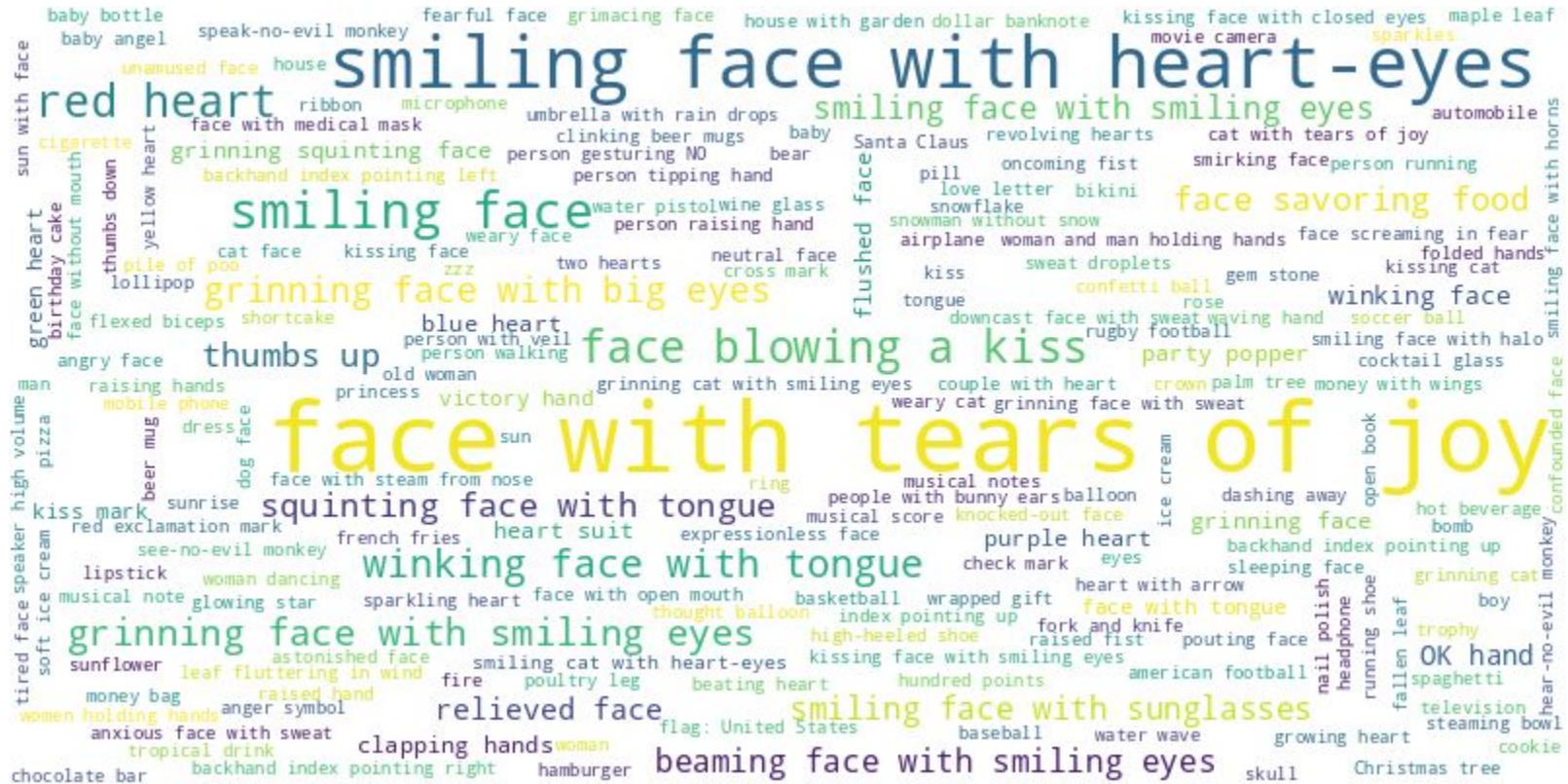
A word cloud visualization centered around the word "Joy". The most prominent word is "Joy" itself, located at the top center. Other large, bold words include "get", "good", "laugh", "LOVE", and "make". The words are rendered in a variety of colors (blue, green, red, orange, yellow, purple) and sizes, creating a dynamic and expressive layout. The words are arranged in a non-linear, overlapping fashion, with some words appearing multiple times in different sizes and colors. The overall theme of the word cloud is positive, joyful, and social.

pretty funny house
say even play go
do way beautiful
yay big loud
ya hair never favorite
someone meet thing
world ill glad hahaha
girl christmas w call
soo right sleep
rt day yes god
want as today
wake bad tell
night much already
cute break
something little
miss start
goodnight
funny
house
even
play
mean
okay
picture
everyone
really
maybe
stay
text
friend
take
morning
birthday
cause
game
school
real
first
movie
oh
show
sweet
look
long
stop
lot
boyfriend
time
ask
excite
ever
gonna
drink
ok
bitch
hope
wish
mom
wanna
know
try
wait
saw
end
old
year
party
everything
life
happy
ready
bed
keep
gotta
yeah
feel
long
stop
well
back
tho
eat
amaze
live
nice
great
sunday
new
room
ask
excite
ever
gonna
drink
ok
bitch
hope
wish
mom
wanna
know
try
wait
saw
end
old
year
party
everything
life
happy
ready
bed
keep
gotta
yeah
feel
long
stop
well
back
tho
eat
amaze
live
nice
great
sunday
new
room
ask
excite
ever
gonna
drink
ok
bitch
hope
wish
mom
wanna
know
try
wait
saw
end
old
year
party
everything
life
happy
ready
bed
keep
gotta
yeah
feel
long
stop
well
back
tho
eat
amaze
live
nice
great
sunday
new
room
ask
excite
ever
gonna
drink
ok
bitch
hope
wish
mom
wanna
know
try
wait
saw
end
old
year
party
everything
life
happy
ready
bed
keep
gotta
yeah
feel
long
stop
well
back
tho
eat
amaze
live
nice
great
sunday
new
room

Wordcloud - Joy (hashtag)



Wordcloud - Joy (emoji)

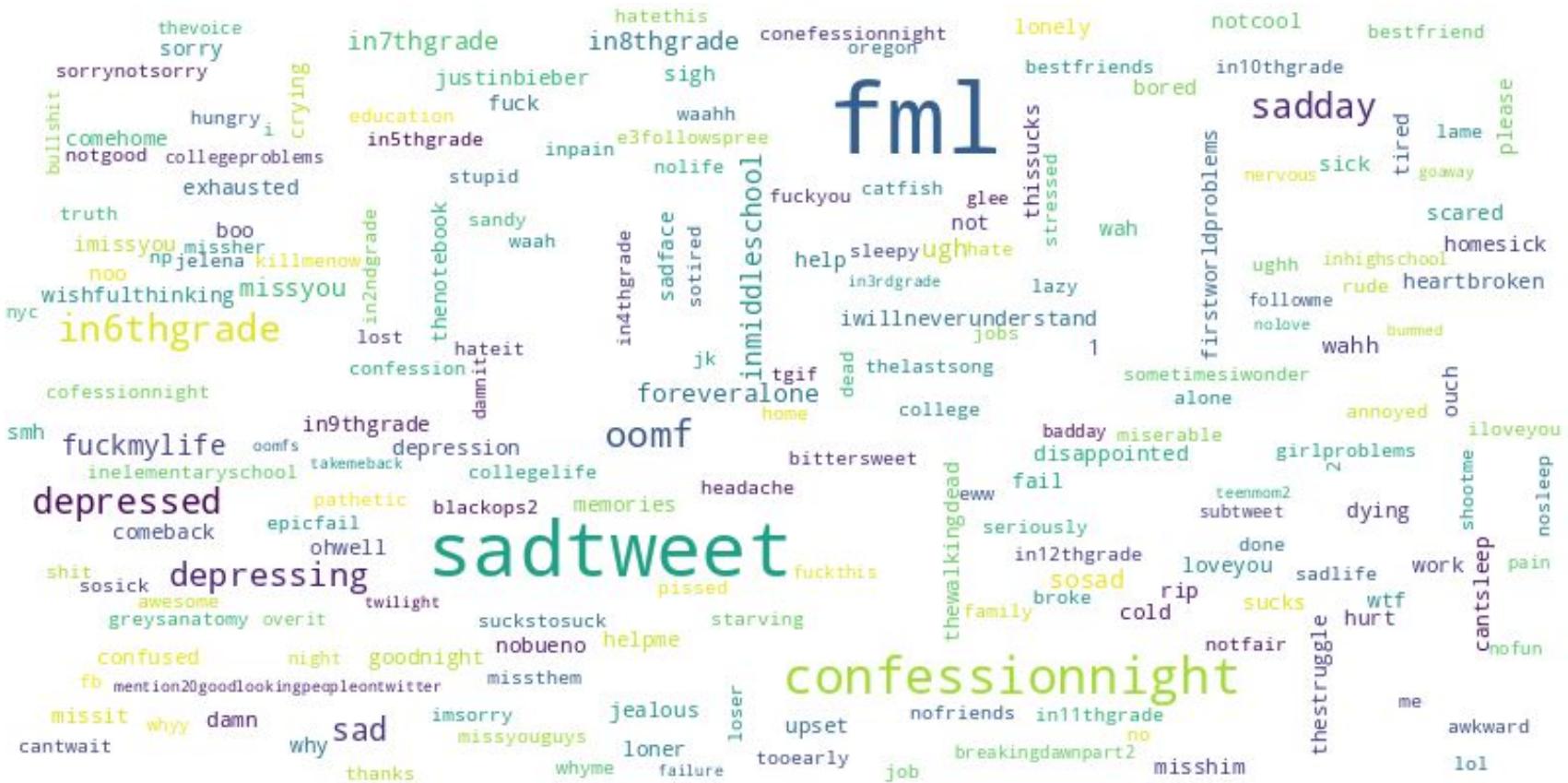


Wordcloud - Joy (emoticon)



Wordcloud - Sadness (word)

Wordcloud - Sadness (hashtag)



Wordcloud - Sadness (emoji)

Wordcloud - Sadness (emoticon)

20

10

:-)

:c

:I

:/

: (

5

: ' (

5

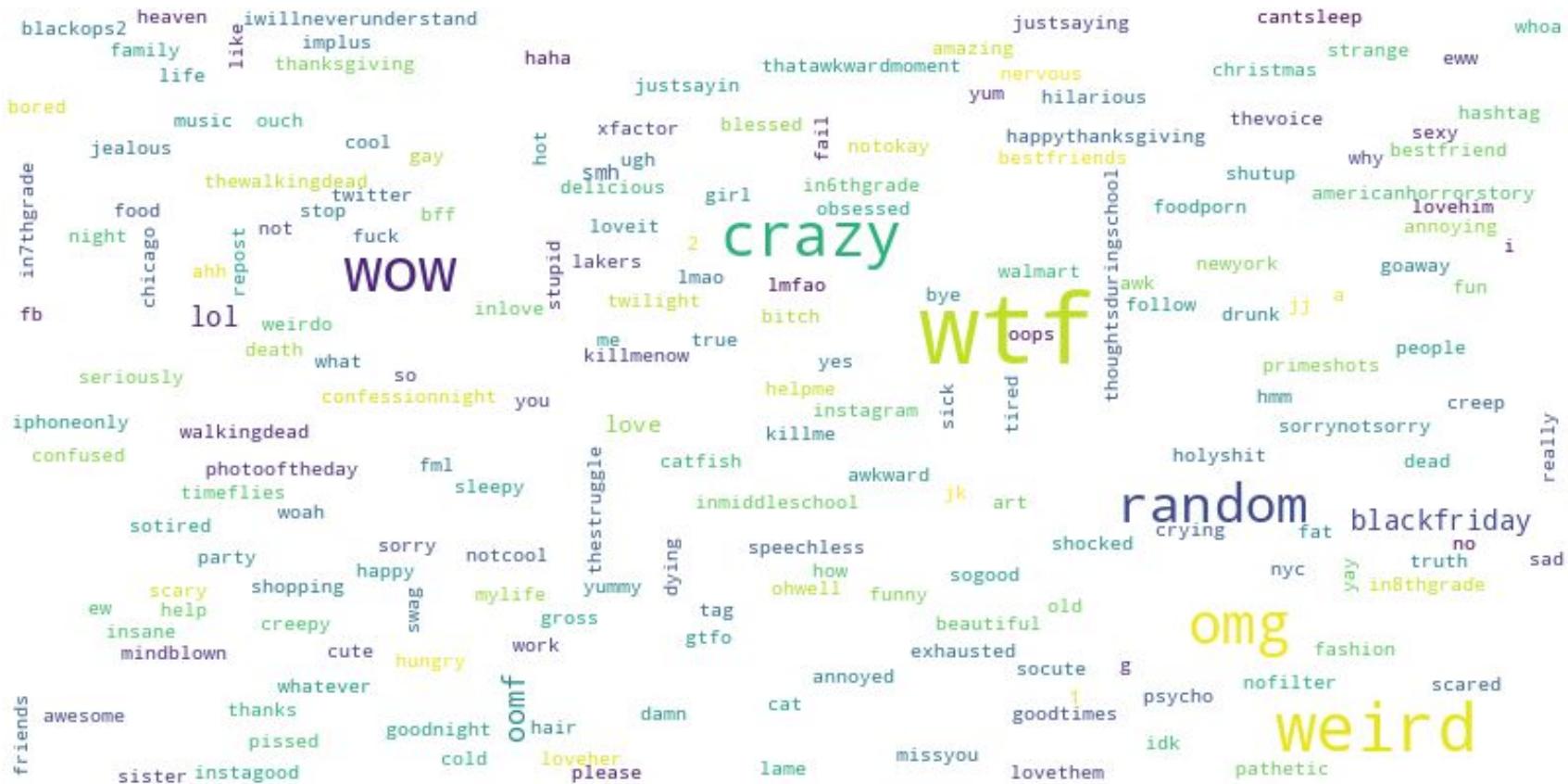
Wordcloud - Surprise (word)

A wordcloud visualization showing the frequency of various words in a surprise text. The words are rendered in different colors and sizes, with larger and more prominent words like 'get', 'laugh', 'loud', 'think', and 'right' appearing multiple times.

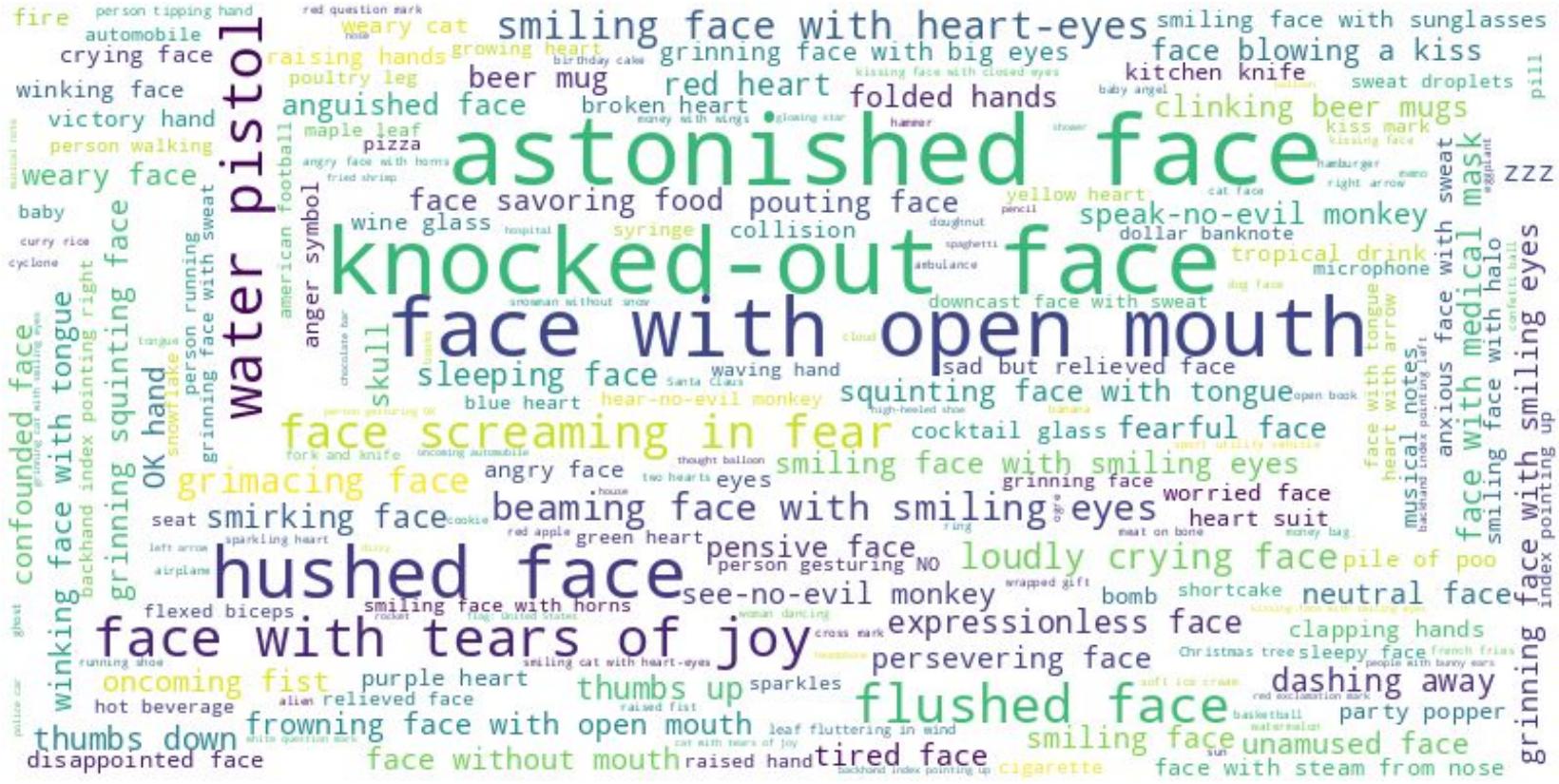
The wordcloud includes the following words:

- follow
- believe
- never
- night
- pretty
- want
- sister
- tonight
- kid
- moment
- wrong
- school
- hope
- damn
- hahaha
- playhouse
- class
- tweet
- oh
- dont
- run
- take
- wait
- man
- remember
- nigga
- something
- god
- high
- bed
- buy
- leave
- guess
- still
- know
- people
- room
- feel
- music
- black
- cause
- see
- talk
- phone
- hair
- face
- thing
- sawfood
- eye
- keep
- happy
- old
- boy
- stop
- try
- smh
- find
- do
- funny
- shit
- bitch
- lmfaot
- rt
- name
- day
- gotta
- long
- way
- call
- go
- use
- sleep
- world
- soo
- much
- car
- time
- awkward
- always
- maybe
- Show
- tomorrow
- first
- happen
- random
- guy
- life
- kill
- work
- crazy
- mean
- hate
- almost
- everyone
- today
- yet
- hear
- dad
- even
- let
- last
- eat
- lot
- miss
- walk
- brother
- dude
- turn
- yes
- live
- next
- whole
- thanksgiving
- big
- weird
- need
- year
- sit
- sit
- come
- real
- ever
- movie
- turn
- seriously
- f
- text
- die
- end
- wonder
- look
- home
- dream
- put
- many
- head
- twitter
- friend
- game
- good
- haha
- wanna
- well
- already
- think
- break
- make
- lose
- give
- picture
- girl
- watch
- family
- new
- month
- as
- okay
- someone
- baby
- ask
- wish
- friday
- gonna
- wear
- sound
- wake
- yeah
- tho
- christmas
- dead
- sure

Wordcloud - Surprise (hashtag)



Wordcloud - Surprise (emoji)



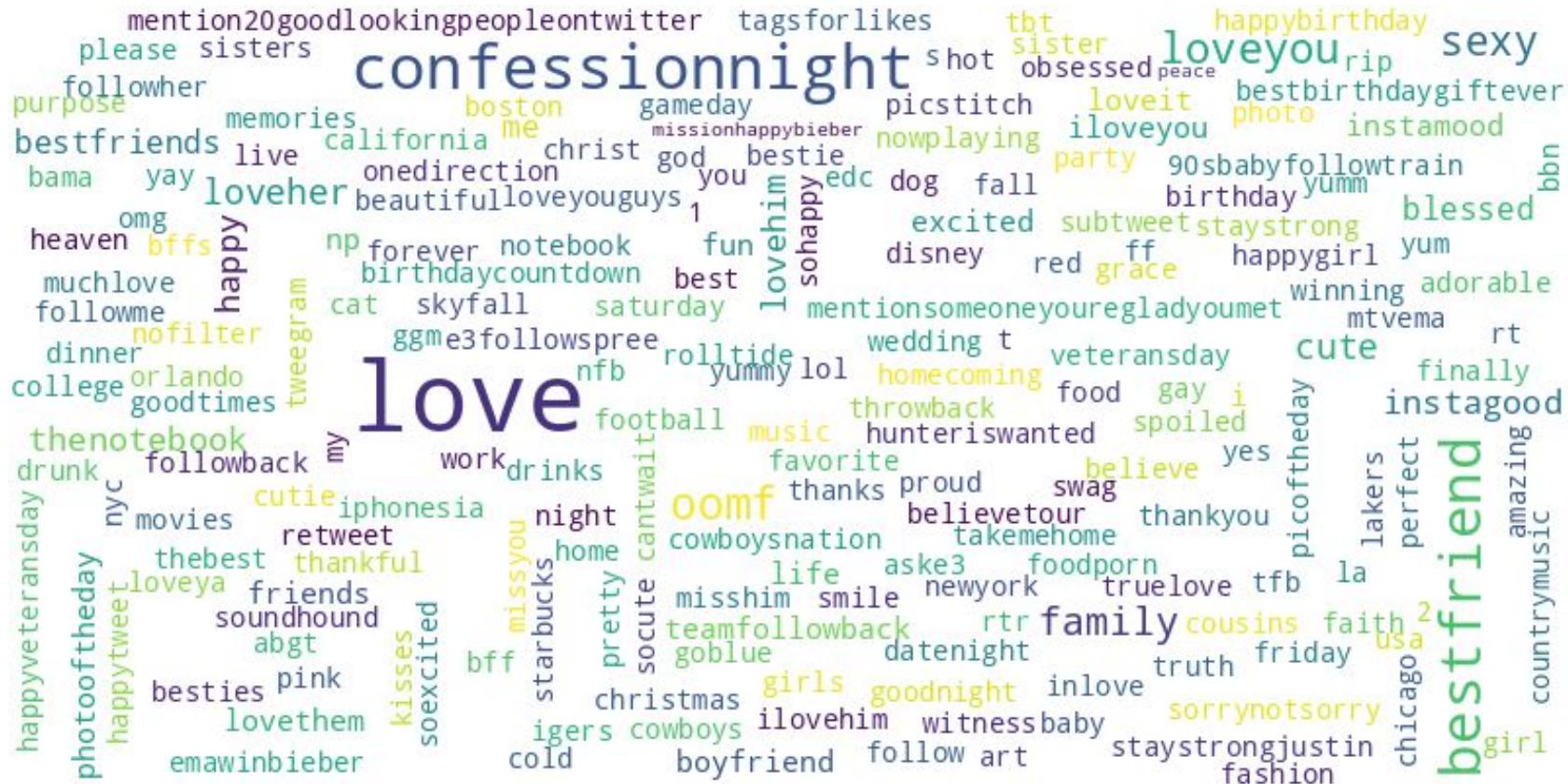
Wordcloud - Surprise (emoticon)



Wordcloud - Trust (word)

cause meet seriously bring music dinner sure sleep little
stop game put people daycall follow home listen bitch kid wanna
long house date care notebook still person sister rt
wish amaze favorite look eat tomorrow smile hot wanna
world bad perfect big hard family hahaha laugh last
thing morning go text aww soon miss well nothing something W
way birthday please right friend best first dont today gotta
picture thank stay together football ask happy finally
try let whole find really fun always mine baby play keep haha
end oh back start sweet forever tell sexy ill never break
bestfriend okay work great next boyfriend hope excite bed
say live movie crazy hey week hour life good feel mean tho
place old year night yeah want fall mom luck talk as tho
babe loud show ya night anything boy buy shit sorry
already know show ya guy ever give think soo real nice
christmashead school get someone you face even glad take fuck
n

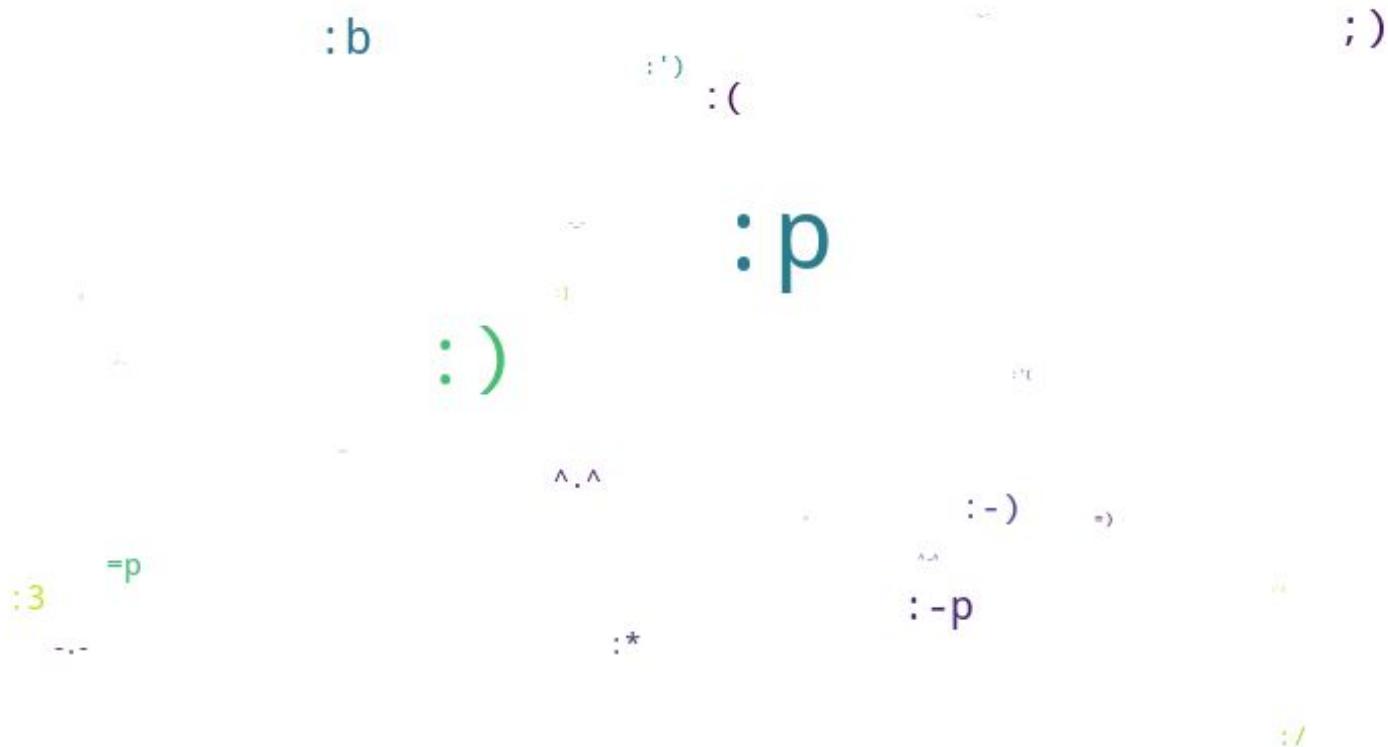
Wordcloud - Trust (hashtag)



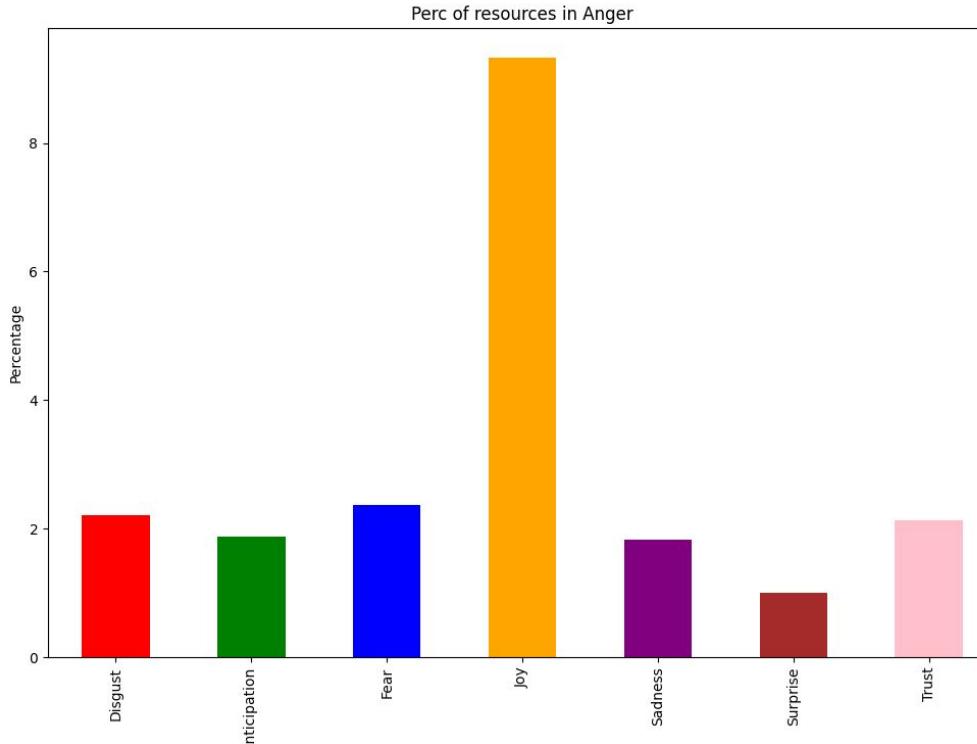
Wordcloud - Trust (emoji)



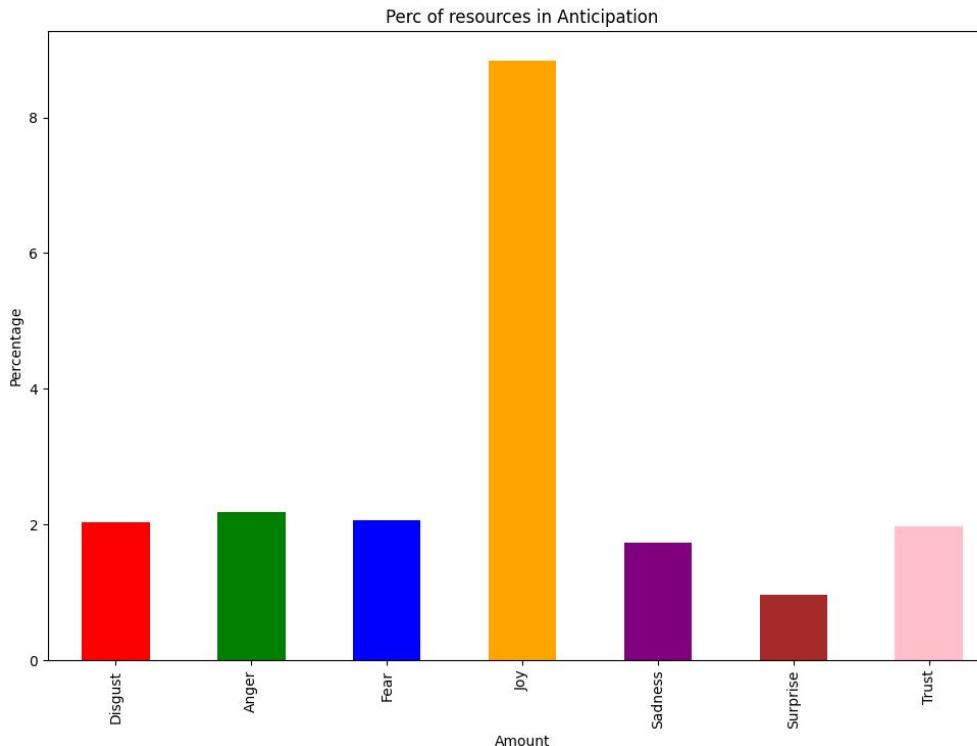
Wordcloud - Trust (emoticon)



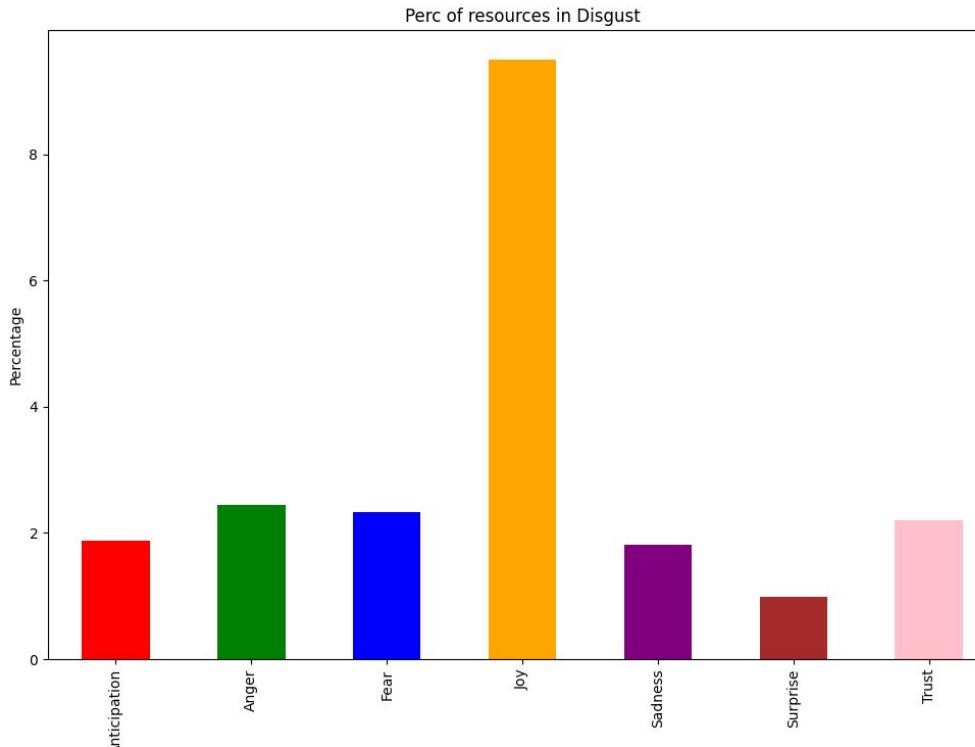
Iistogramma - Anger



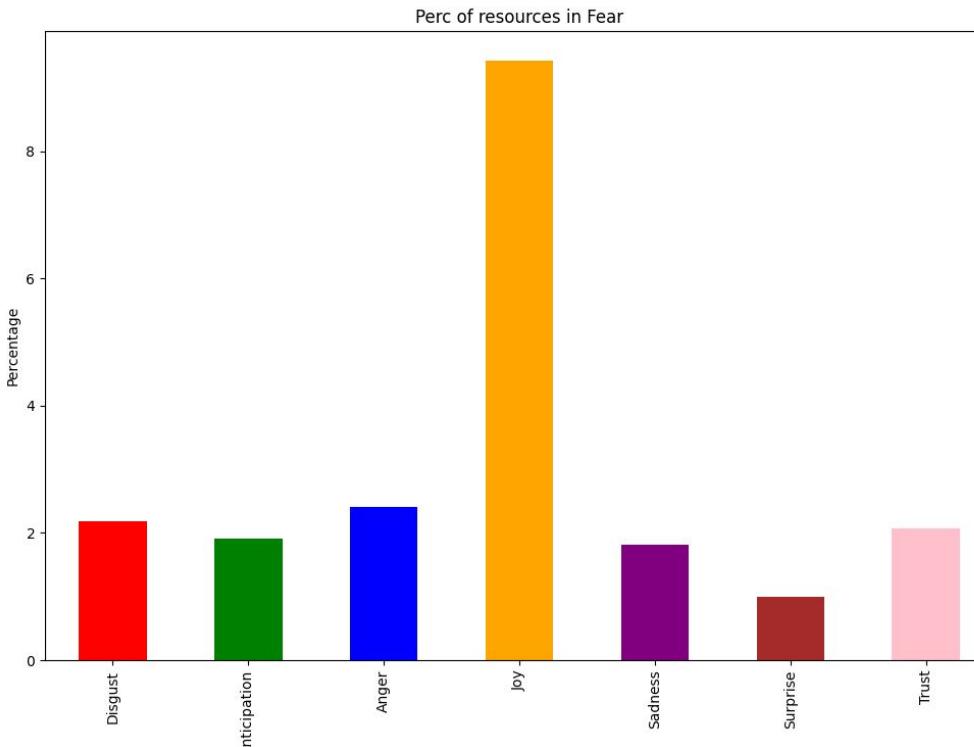
Iistogramma - Anticipation



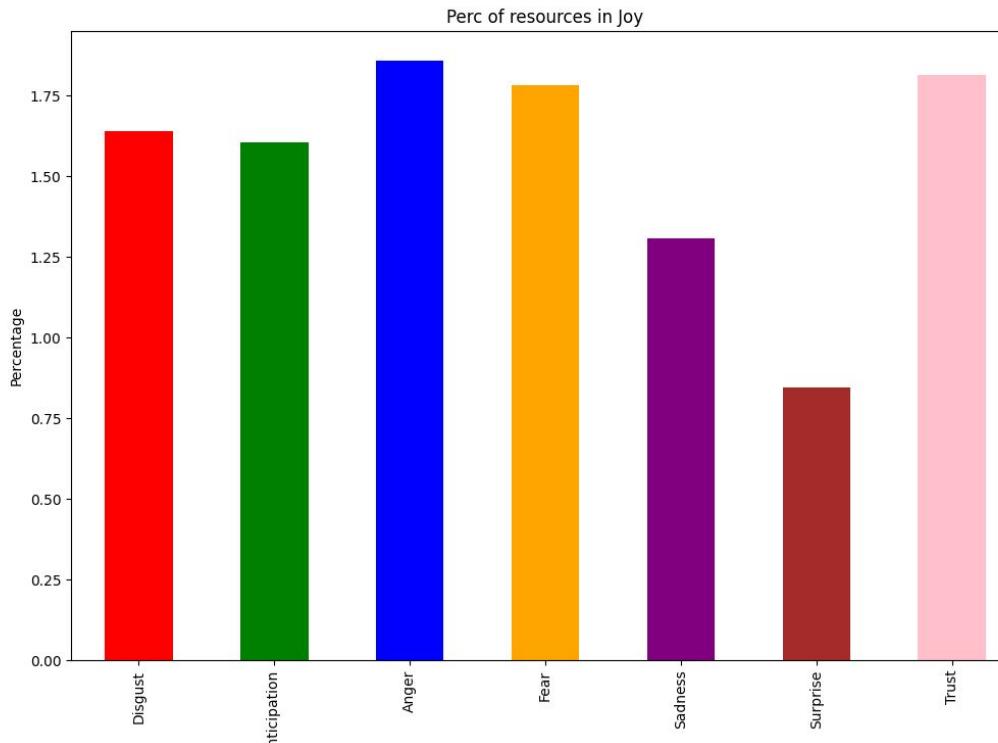
Iistogramma - Disgust



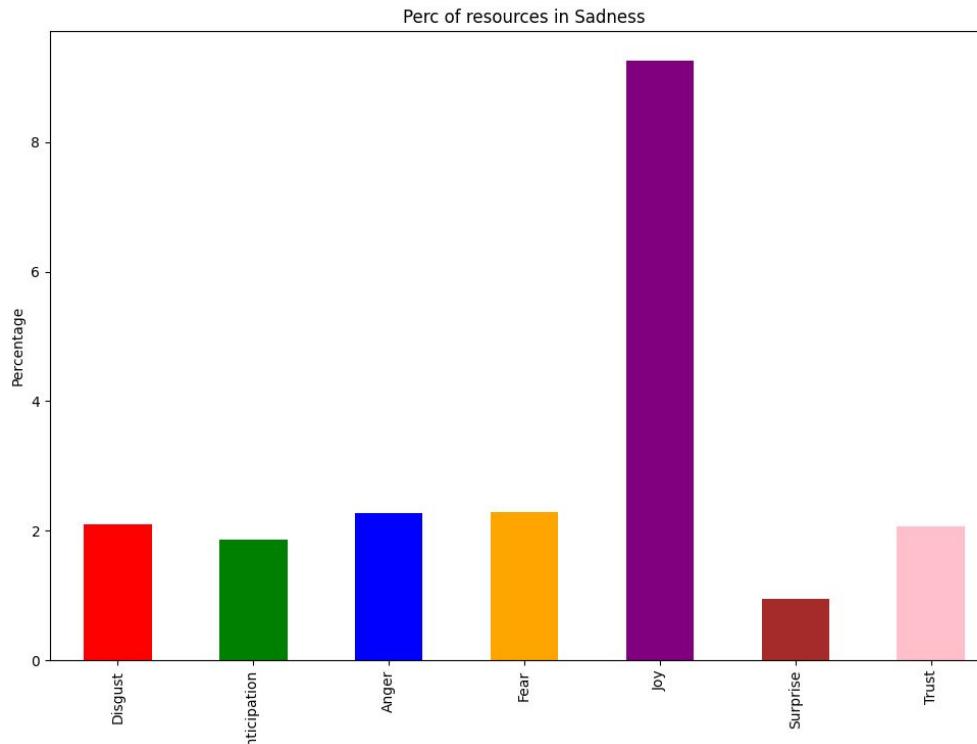
Iistogramma - Fear



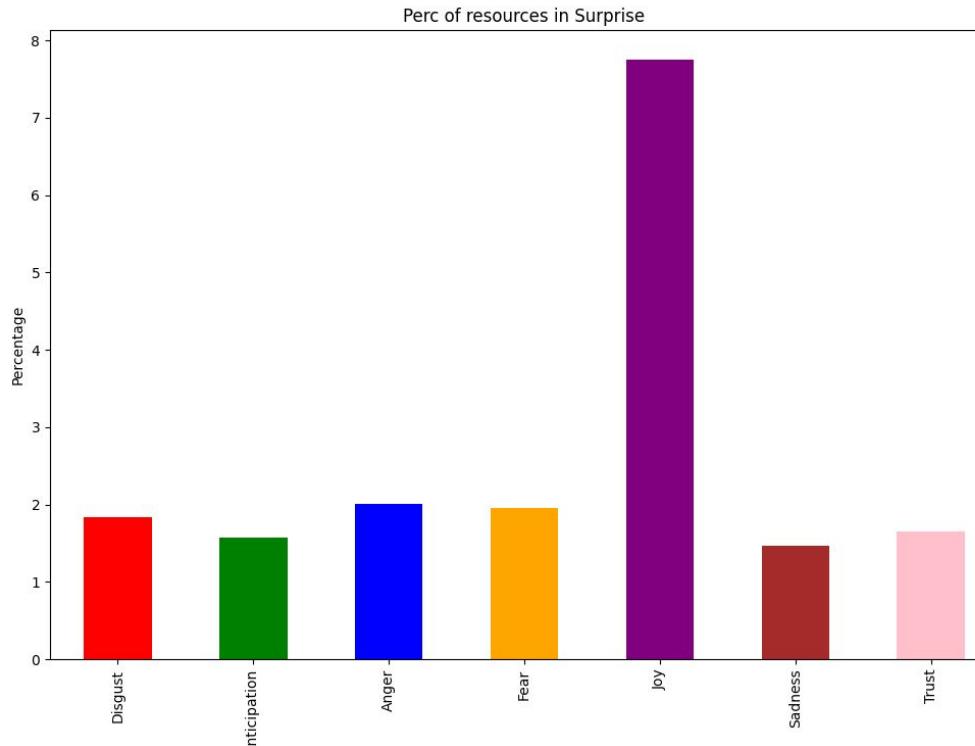
Iistogramma - Joy



Iistogramma - Sadness



Iistogramma - Surprise



Iistogramma - Trust

