MAADB

Obiettivi

1. Costruire una word cloud delle parole con la frequenza maggiore per ognuna delle emozioni
2. Fare statistiche su hashtag ed emoticon
3. Arricchire il tesauro con le parole più frequenti nei tweet ma non presenti in esso ed eliminare da quest’ultimo quelle con la frequenza bassa (approfondire la scelta del valore soglia).
4. Dire per ogni risorsa, quali parole compaiono nei tweet e quali no. Inoltre dire quali nuove parole non presenti nelle risorse ma presenti nei messaggi sono state trovate
5. Comparare i risultati tra il nuovo tesauro creato e il vecchio
6. Comparare le performance tra dbms relazionale e nosql
7. Proporre una soluzione di machine learning su sentiment analisys

La gestione delle negazioni avviene considerando la parola come se fosse preceduta da un not.

Nella word cloud avremo ad esempio “not good”

Possibili estensioni:

considerare le stop words, piuttosto che in maniera statica, con un procedimento dinamico che capisce da solo quali sono le stop words.

Stessa cosa per emoji/emoticons e slang words

Linguaggio: Python

DBMS: MongoDB e Microsoft swl server (Express Edition).

Per MongoDB 4 nodi da mandare in esecuzione 2 su un pc e 2 sull’altro.

Creare in python un sistema a web services. Ci sarà un programma principale che chiama i programmi con i servizi web per ottenere i dati da ridurre.

SCHEMA DEI DATI

**Tabella conteggi parole** (conterrà il risultato della funzione map) IN MONGO DB FARLO INNESTATO

Parola, sentimento, flagSS, flagNrc, flagEMO, numeroDiOccorrenzeNelleFrasi

**Tabella conteggi emoticon/emoji** INNESTATO

Codice, sentimento, numeroDiOccorrenzeNelleFrasi

**Tabella conteggi hashtag** INNESTATO

Hashtag, sentimento, numeroDiOccorrenzeNelleFrasi

Tabella delle frasi

ID, sentimento, frase

Tabella stopwords (STATICA)

parola

Tabella negative words (STATICA)

parola

Tabella slang (STATICA)

Slang, traduzione

Tabella emoticon/emoji (STATICA)

ID,Codice, polarità (1 positivo, 0 neutro, -1 negativo)

DA FARE:

Angelo schema in mongodb

Amedeo schema in sqlserver

Implementare infrastruttura generale basata su web services

Implementare caricamento dei dati generali (risorse, emoticons, ecc…)

Parsing…….