

Sentiment Analysis

Data Wrangling



Agenda

- Obtener data del API de twitter utilizando R.
- Generar un corpus.
- Sentiment Analysis:
 - Subjetividad
 - Polaridad
- Lematización

Vector vs Corpus

VectorSource

Content_1

Content_2

Content_3

VCorpus

Doc_1

Content_1

Meta

Doc_2

Content_2

Meta

Doc_3

Content_3

Meta

Ejemplo

```
[1] "@HeatherWhaley I was about 2 joke it takes 2 hands to hold hot  
coffee...then I read headline! #Don'tDrinkNShoot"
```

```
$meta
```

```
author      : character(0)  
timestamp   : 2019-08-07 00:25:22  
description : character(0)  
heading     : character(0)  
id          : 15  
language    : en  
origin      : character(0)  
,
```

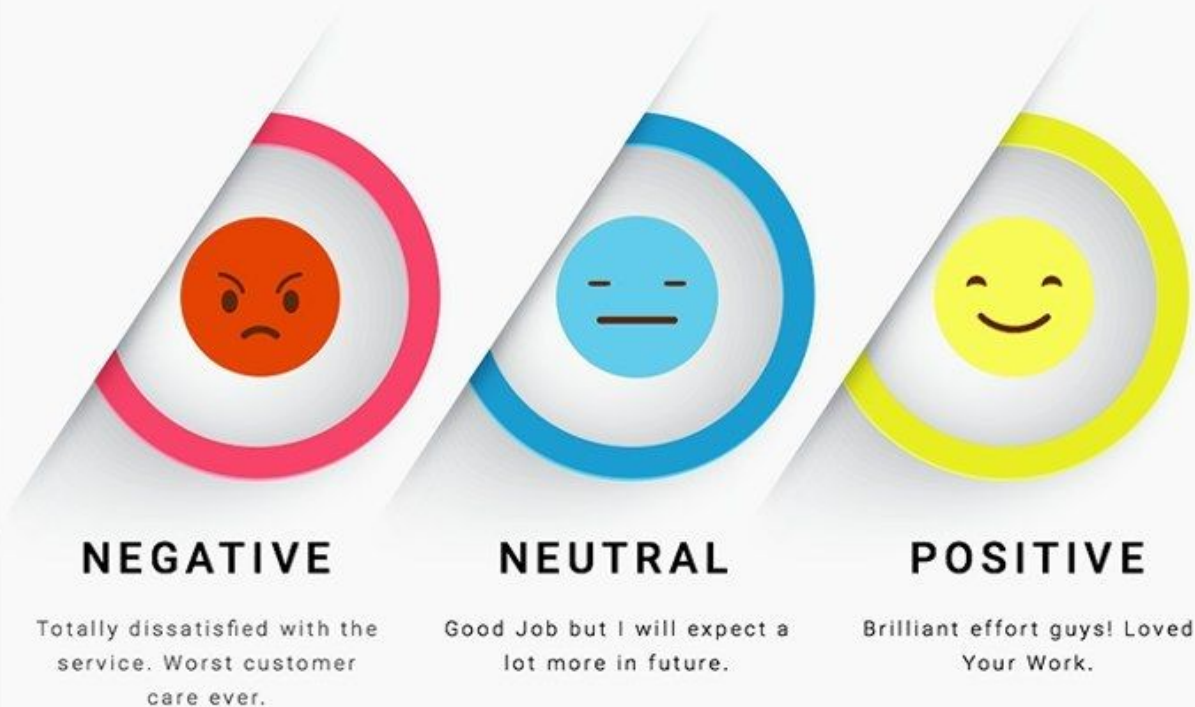
Clasificación de Texto

Polaridad

Queremos saber si la opinión expresada en un documento es positiva, negativa o neutra.

Rango: [-1,1]

SENTIMENT ANALYSIS



Subjetividad

Analizar si el enunciado propuesto es objetivo, es decir, no expresa sentimientos explícitos, o subjetivo, expresa sentimientos.

Rango: [0,1]

Una reseña de una película o restaurante tiene cierto valor de subjetividad.

Una oración objetiva presenta hechos de manera concisa y concreta. (Valor cercano a 0)

”Joker es la mejor película del año y merece ganar el Oscar.”

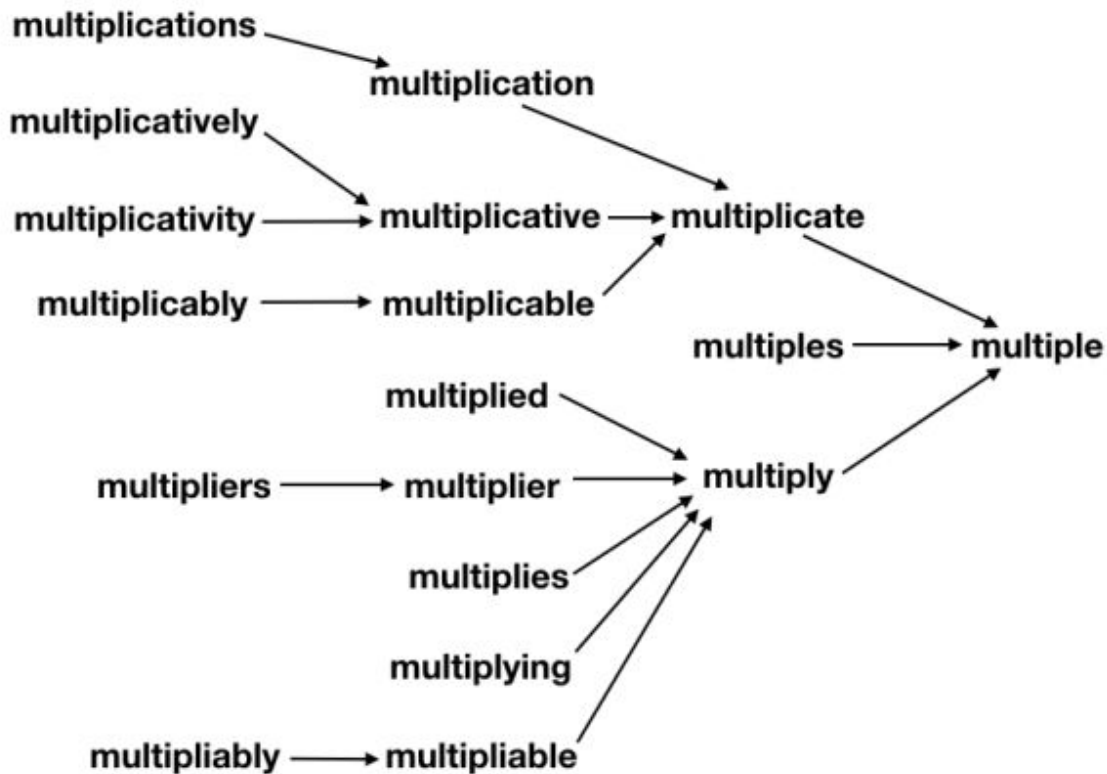
Joaquin Phoenix protagoniza el papel de Joker.

Lematización

Método que reduce la inflexión verbal de las palabras a la raíz de su lenguaje.

A la raíz se le conoce como lemma.

Lemma: Forma canónica o de diccionario de un conjunto de palabras.



Lematización

Método que reduce la inflexión verbal de las palabras a la raíz de su lenguaje.

A la raíz se le conoce como lemma.

Lemma: Forma canónica o de diccionario de un conjunto de palabras.

	original_word	lemmatized_word
0	trouble	trouble
1	troubling	trouble
2	troubled	trouble
3	troubles	trouble

	original_word	lemmatized_word
0	goose	goose
1	geese	goose