




# Analizador de texto

Juan Manuel Restrepo Urrego  
Stiven Valencia Ramírez



Antes de empezar a mostrar el proyecto, tenemos que decir que empezamos con textos cortos, osea oraciones, luego parrafos y por ultimo textos.

Aunque también probamos con varios textos a la vez.

De esta manera el proyecto fue creciendo de forma correcta y ordenada sabiendo que se estaba haciendo poco a poco ahora algunas pruebas

# Sentimientos

En este caso se probó con texto en inglés para facilitar a la librería el reconocimiento de palabras, como vemos el analizador de sentimientos separa el texto por los puntos y analiza cada oración por aparte diciéndonos los indicadores de negatividad, positividad o neutralidad y por último nos dice si fue bueno o malo el texto con un compendio

```
tokenizer = nltk.data.load('tokenizers/punkt/english.pickle')
sentences = tokenizer.tokenize("I am sure that is the reason why education is so important. When you are truly interested in something, you never stop learning. I think this taste awful, I dont like it.")
```

```
I am sure that is the reason why education is so important.
neg : 0.0

neu : 0.659

pos : 0.341

compound : 0.567

When you are truly interested in something, you never stop learning.
neg : 0.0

neu : 0.517

pos : 0.483

compound : 0.7571

I think this taste awful, I dont like it.
neg : 0.505

neu : 0.495

pos : 0.0

compound : -0.6261
```

# Nombres o personajes importantes

En este caso no tenemos muchos problemas ya que con nltk que es una herramienta de lenguaje natural nos dice después de analizar y tokenizar un texto si las palabras son pronombres, sujetos, verbos, etc.

Entonces primero depuramos el texto y por último damos los resultados para mayor eficiencia y velocidad

```
import nltk
from nltk import word_tokenize
s="This ins't a very long sentence but it's full of interesting words, pablo"
tags = nltk.pos_tag(word_tokenize(s))
nouns = [word for word, tag in tags if tag in ["NN", "NNS", "NNP", "NNPS"]]
tags
```

```
[('This', 'DT'),
 ("ins't", 'VBZ'),
 ('a', 'DT'),
 ('very', 'RB'),
 ('long', 'JJ'),
 ('sentence', 'NN'),
 ('but', 'CC'),
 ('it', 'PRP'),
 ('s', 'VBZ'),
 ('full', 'JJ'),
 ('of', 'IN'),
 ('interesting', 'JJ'),
 ('words', 'NNS'),
 (',', ','),
 ('pablo', 'NN')]

[word for (word, tag) in tags if "NN" in tag]


['sentence', 'words', 'pablo']
```



# Tema del texto

Este fue el más complicado de todos por su nivel de escala, ya que no es lo mismo analizar partes de un texto a analizar todo un texto de forma inmediata y decir este es el tema principal, ya que antes se debe empezar por un proceso de depuración y limpieza además de dejar palabras claves.

Empezamos con oraciones, luego con párrafos y por último el texto completo aquí los resultados



Ahora como código final probamos este texto y estos fueron los resultados en el caso de sentimientos, recordar que el compendio es el importante en este caso si el número de compendio es negativo pues es malo y al contrario con el positivo y si se aproxima a cero es neutro.

cmd: Símbolo del sistema

```
C:\Users\stivenvr\Desktop>python CompuBlanda.py
---LAS 10 PALABRAS MAS USADAS EN EL TEXTO---
christmas
traditions
percent
festive
brits
revealed
"
day
survey
found
```

```
---SENTIMIENTO ENCONTRADO EN EL TEXTO---
0.4523
```

```
---SUJETOS ENCONTRADOS EN EL TEXTO---
```

```
percent
traditions
survey
poll
adults
mince
impact
pandemic
pantomime
christmas
love
bid
children
```

```
C:\Users\stivenvr\Desktop>_
```


cmd: Seleccionar Símbolo del sistema

```
C:\Users\stivenvr\Desktop>python CompuBlanda.py
---LAS 10 PALABRAS MAS USADAS EN EL TEXTO---
drug
treatment
study
's
adults
paxlovid
fda
given
people
placebo
```

```
---SENTIMIENTO ENCONTRADO EN EL TEXTO---
-0.072340625
```

```
---SUJETOS ENCONTRADOS EN EL TEXTO---
```

```
drug
study
company
treatment
month
risk
deaths
data
statement
virus
placebo
fda
end
news
pills
name
```



Para nombres o personajes importantes tenemos que se usó los sustantivos de las diferentes divisiones de oraciones del texto inicial y aquí los resultados

```
name  
hospital  
paxlovid  
adults  
efficacy  
result  
research  
live  
world  
bourla  
omicron  
need  
pandemic  
enzyme  
spike  
mutations  
date  
committee  
mix  
group  
symptoms  
endpoint  
majority  
reason  
year  
time  
day  
pipeline  
analysis  
brand  
pill
```

```
C:\Users\stivenvr\Desktop>_
```