

## Proyecto Análisis de Sentimientos

El objetivo de este proyecto es diseñar un algoritmo que evalúe qué tan positivo, neutral o negativo es una cierta frase.

Para esto vamos a buscar ciertas palabras clave específicas y asignarles una valoración. Como ejemplo, consideremos la frase

“Excelente en su área, su muerte es una enorme pérdida y debería ser luto nacional”

Supongamos también que hemos elegido como palabras clave las palabras *muerte*, *pérdida*, *luto*, *excelente*, *gran* y *positivo*. Para construir un modelo vectorial para analizar esta palabra, construiremos un vector de unos y ceros que indique si las palabras clave están o no en la frase a estudiar. Para nuestro ejemplo, el vector quedaría

$$w = (1, 1, 1, 1, 0, 0)$$

porque las primeras cuatro palabras clave se encuentran en la frase a estudiar pero las últimas dos no. A continuación construiremos un vector que cuente la cantidad de palabras positivas, neutras y negativas que tenemos en la frase. Supongamos que para nuestro ejemplo, hemos decidido considerar las palabras

- *excelente*, *gran* y *positivo* como positivas.
- *pérdida* como neutral.
- *muerte* y *luto* como negativa.

Construiremos un vector que cuente la cantidad de palabras positivas, neutras y negativas que se encuentren en la frase a estudiar. En nuestro ejemplo, este vector tendría la forma

$$s = (1, 1, 2)$$

porque tenemos una palabra positiva, una palabra neutra, y dos palabras negativas.

Introducimos ahora las siguientes definiciones:

1. La calidad promedio de una frase se define como la suma de las coordenadas del vector  $w$  dividido la cantidad de palabras clave a estudiar. La denotaremos por  $\text{avg}(w)$ .
2. El promedio del sentimiento de cada frase se define como  $(1, 0, -1) \cdot s$ .

### Se pide:

1. Elija un conjunto de 15 frases a estudiar. Por ejemplo, pueden tomarse publicaciones de alguna red social.
2. En base a las frases elegidas, elabore una lista de palabras clave que considere adecuadas, y clasifíquelas en positivas, neutrales y negativas.
3. Hallar los vectores  $w$  y  $s$  para cada frase.
4. Hallar calidad promedio y el promedio de sentimiento de cada frase. A continuación se muestra un ejemplo de como presentar estos resultados.

	<b>Tweet</b>	<b>Calidad</b>	<b>positiva</b>	<b>neutra</b>	<b>negativa</b>
<b>0</b>	Excelente en su área, su muerte es una enorme ...	0.294118	0.4	0.0	0.6
<b>1</b>	Vaya señora que bueno que se asesora por algui...	0.117647	1.0	0.0	0.0
<b>2</b>	Se me ocurre y sin ir a clase me informéis por...	0.058824	1.0	0.0	0.0
<b>3</b>	Soy docente universitario, estoy intentando pr...	0.352941	0.4	0.6	0.0

5. ¿Cuál frase es más positiva?
6. ¿Cuál frase es más negativa?
7. Explique con sus palabras qué representa la calidad promedio que hemos definido y por qué tiene sentido la definición matemática proporcionada.
8. Explique con sus palabras qué representa el promedio de sentimiento que hemos definido y por qué tiene sentido la definición matemática proporcionada.
9. Explique con sus palabras cómo afecta a este algoritmo, la elección inicial de las palabras clave.
10. Investigue y proponga posibles mejoras para este algoritmo.

### Sobre el informe:

- El tiempo para entregar el informe es hasta el sábado 31 de agosto inclusive. La entrega se realizará por webasignatura.
- El informe deberá estar en formato pdf, la entrega también deberá incluir el proyecto compartido para su fácil ejecución y testeo.
- El informe deberá contener título, fecha, nombre y cédula del estudiante.
- Se evaluará: prolijidad del informe, utilización correcta del idioma español, redacción, prolijidad del código presentado en los scripts, conclusiones.