

# Anexo 20

## Proyecto 20: Manejo de TensorFlow para NLP

---

Mg. Luis Felipe Bustamante Narváez

En este ejercicio, aplicaremos sentencias de TensorFlow para procesar estructuras de datos, con el fin de prepararnos para siguientes proyectos.

### Librerías

```
In [...] from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
```

### Datos

```
In [...] oraciones = ['me gusta el fútbol e ir al estadio',
                    'juego fútbol los fines de semana.',
                    'no me gusta perder']
```

### Procesamiento de Datos

```
In [...] #Vocabulario máximo
max_vocab_size = 30000
#Iniciamos el tokenizador
tokenizer = Tokenizer(num_words=max_vocab_size)
#Tokenizamos
tokenizer.fit_on_texts(oraciones)
#Creamos las secuencias
secuencias = tokenizer.texts_to_sequences(oraciones)
```

```
In [...] print(secuencias)

[[1, 2, 4, 3, 5, 6, 7, 8], [9, 3, 10, 11, 12, 13], [14, 1, 2, 15]]
```

```
In [...] #Diccionario de palabras (los ordena por peso: mayor frecuencia)
tokenizer.word_index
```

```
Out[... {'me': 1,
        'gusta': 2,
        'fútbol': 3,
        'el': 4,
        'e': 5,
        'ir': 6,
        'al': 7,
        'estadio': 8,
        'juego': 9,
        'los': 10,
        'fines': 11,
        'de': 12,
        'semana': 13,
        'no': 14,
        'perder': 15}
```

```
In [... #Organizamos la secuencia estandarizando la cantidad de elementos por vec
data = pad_sequences(secuencias)
print(data)
```

```
[[ 1  2  4  3  5  6  7  8]
 [ 0  0  9  3 10 11 12 13]
 [ 0  0  0  0 14  1  2 15]]
```

```
In [... #Limitamos el número de palabras con relleno al ppio
max_sequence_length = 5
data = pad_sequences(secuencias, maxlen=max_sequence_length)
print(data)
```

```
[[ 3  5  6  7  8]
 [ 3 10 11 12 13]
 [ 0 14  1  2 15]]
```

```
In [... #Limitamos el número de palabras con relleno al final
max_sequence_length = 5
data = pad_sequences(secuencias, maxlen=max_sequence_length, padding='pos'
print(data)
```

```
[[ 3  5  6  7  8]
 [ 3 10 11 12 13]
 [14  1  2 15  0]]
```

```
In [... # cambiamos los valores directamente en el método
data = pad_sequences(secuencias, maxlen=6)
print(data)
```

```
[[ 4  3  5  6  7  8]
 [ 9  3 10 11 12 13]
 [ 0  0 14  1  2 15]]
```

# Conclusiones

Este proyecto permite indagar sobre algunas sentencias de TensorFlow para procesar datos, tema que será de gran aporte, cuando desarrollemos proyectos robustos de Inteligencia Artificial.

---

Mg. Luis Felipe Bustamante Narváez