

Python

Procesamiento de Lenguaje Natural

Diversidad Léxica



Rogelio Ferreira Escutia

Profesor / Investigador
Tecnológico Nacional de México
Campus Morelia



Diversidad Léxica

Diversidad Léxica

- La “Diversidad Léxica” es un parámetro que nos indica la “riqueza” de un texto, es decir, la forma de expresión de un texto, donde por lo general, una mayor cantidad de palabras utilizadas en un mismo texto, nos lleva a considerar que tiene mas “riqueza” léxica.
- Hay varias formas de calcular la “Diversidad Léxica” (la cual ya lleva muchos años siendo estudiada), pero para este ejemplo utilizaremos una fórmula muy sencilla:

Diversidad Léxica = cantidad de palabras / cantidad de caracteres

Diversidad (1)

- **Biblioteca a utilizar:**

```
# Importar bibliotecas  
import nltk
```

```
# Procesamiento del lenguaje natural
```

Diversidad (2)

- **Definir texto:**

```
# Definir una frase
frase_amado_nervo = "Amé, fui amado, el sol acarició mi faz. ¡Vida, nada me debes! ¡Vida, estamos en paz!"

# Definir la frase a procesar
texto = frase_amado_nervo

# Impresión en pantalla de la frase seleccionada
print("\nTexto Completo:\n", texto)
```

- **Salida:**

```
Texto Completo:
Amé, fui amado, el sol acarició mi faz. ¡Vida, nada me debes! ¡Vida, estamos en paz!
```

Diversidad (3)

- **Calcular caracteres y palabras:**

```
# Cantidad de caracteres
caracteres = len(texto)
print("\nCantidad de caracteres: ", caracteres)

# Cantidad de palabras
palabras = len(set(texto))
print("Cantidad de palabras: ", palabras)
```

- **Salida:**

```
Cantidad de caracteres: 84
Cantidad de palabras: 26
```

Diversidad (4)

- **Calculando la Diversidad Léxica:**

```
# Diversidad Léxica
diversidad = palabras / caracteres
print("\nDiversidad Léxica: ", diversidad, "\n")
```

- **Salida:**

```
Diversidad Léxica: 0.30952380952380953
```



Rogelio Ferreira Escutia

Profesor / Investigador
Tecnológico Nacional de México
Campus Morelia



rogelio.fe@morelia.tecnm.mx



rogeplus@gmail.com



xumarhu.net



[@rogeplus](https://twitter.com/rogeplus)



[https://www.youtube.com/
channel/UC0on88n3LwTKxJb8T09sGjg](https://www.youtube.com/channel/UC0on88n3LwTKxJb8T09sGjg)



[rogelioferreiraescutia](https://www.linkedin.com/in/rogelioferreiraescutia)

