

creaciondiccionario-a01633819

October 18, 2023

0.1 Actividad NLP - 1.0 - Creación de Diccionario

Andrés Alejandro Guzmán González - A01633819

Inteligencia artificial avanzada para la ciencia de datos II

Grupo 502

```
[60]: # Llamado a librerías
import matplotlib.pyplot as plt
from collections import Counter
from wordcloud import WordCloud
```

Función para guardar el contenido del txt como string

```
[61]: def txt_2_str(archivo):
    with open(archivo, 'r') as file:
        contenido_texto = file.read()
    return contenido_texto
```

Función para generar un diccionario con las palabras del texto

```
[62]: def crear_diccionario(texto):
    palabras = texto.split()
    contador = Counter(palabras)
    return dict(contador)
```

Código

```
[63]: '''
    Importamos el archivo txt y se convierte a una cadena de caracteres (string)
    '''
    archivo = '/content/sample_data/archivo.txt'
    data = txt_2_str(archivo)
    data
```

```
[63]: 'La Inteligencia Artificial (IA) es un campo de la informática que se ha
convertido en una de las tecnologías más disruptivas y transformadoras de la era
moderna. Esta disciplina se centra en desarrollar sistemas informáticos capaces
```

de realizar tareas que normalmente requerirían inteligencia humana, como el aprendizaje, la percepción y la toma de decisiones. La IA ha experimentado avances espectaculares en los últimos años, impulsados por el aumento de la potencia computacional y la disponibilidad de grandes cantidades de datos. Esto ha permitido la creación de sistemas de IA que pueden realizar tareas tan diversas como el reconocimiento de voz, la traducción automática, la conducción autónoma y el diagnóstico médico, entre muchas otras.\n\nUno de los aspectos más emocionantes de la IA es su capacidad para aprender y adaptarse. Los sistemas de IA utilizan algoritmos de aprendizaje automático para analizar datos y mejorar su rendimiento con el tiempo. Esto ha dado lugar a aplicaciones como los asistentes virtuales, que pueden comprender el lenguaje natural y responder a preguntas de manera cada vez más precisa. Sin embargo, esta creciente sofisticación plantea cuestiones éticas y de privacidad que deben ser abordadas con cuidado, ya que la IA también puede utilizarse de manera inapropiada o para fines perjudiciales.\n\nA medida que la IA continúa evolucionando, se espera que tenga un impacto significativo en una amplia variedad de industrias, desde la atención médica y la educación hasta la industria automotriz y la logística. También plantea desafíos importantes en términos de regulación y responsabilidad, lo que hace que la ética en la IA sea un tema crucial para su desarrollo futuro. En última instancia, la IA tiene el potencial de mejorar la eficiencia y la calidad de vida, pero también requiere una gestión cuidadosa para garantizar que se utilice de manera beneficiosa y responsable.'

```
[64]: '''
Se genera un diccionario, siendo la clave (Key) la palabra y el segundo dato la
frecuencia de aparición de la palabra en el texto.
'''
diccionario = crear_diccionario(data)
'''
Se ordena el diccionario de mayor a menor de acuerdo a la frecuencia.
'''
diccionario= dict(sorted(diccionario.items(), key=lambda item: item[1],
↪reverse=True))
for palabra, frecuencia in diccionario.items():
    print(f'{palabra}: {frecuencia}')
```

```
de: 23
la: 21
y: 13
que: 11
IA: 8
el: 7
en: 6
para: 5
se: 4
ha: 4
un: 3
```

una: 3
más: 3
sistemas: 3
como: 3
los: 3
su: 3
manera: 3
La: 2
es: 2
realizar: 2
tareas: 2
Esto: 2
pueden: 2
mejorar: 2
con: 2
a: 2
plantea: 2
también: 2
Inteligencia: 1
Artificial: 1
(IA): 1
campo: 1
informática: 1
convertido: 1
las: 1
tecnologías: 1
disruptivas: 1
transformadoras: 1
era: 1
moderna.: 1
Esta: 1
disciplina: 1
centra: 1
desarrollar: 1
informáticos: 1
capaces: 1
normalmente: 1
requerirían: 1
inteligencia: 1
humana,: 1
aprendizaje,: 1
percepción: 1
toma: 1
decisiones.: 1
experimentado: 1
avances: 1
espectaculares: 1
últimos: 1

años,: 1
impulsados: 1
por: 1
aumento: 1
potencia: 1
computacional: 1
disponibilidad: 1
grandes: 1
cantidades: 1
datos.: 1
permitido: 1
creación: 1
tan: 1
diversas: 1
reconocimiento: 1
voz,: 1
traducción: 1
automática,: 1
conducción: 1
autónoma: 1
diagnóstico: 1
médico,: 1
entre: 1
muchas: 1
otras.: 1
Uno: 1
aspectos: 1
emocionantes: 1
capacidad: 1
aprender: 1
adaptarse.: 1
Los: 1
utilizan: 1
algoritmos: 1
aprendizaje: 1
automático: 1
analizar: 1
datos: 1
rendimiento: 1
tiempo.: 1
dado: 1
lugar: 1
aplicaciones: 1
asistentes: 1
virtuales,: 1
comprender: 1
lenguaje: 1
natural: 1

responder: 1
preguntas: 1
cada: 1
vez: 1
precisa.: 1
Sin: 1
embargo,: 1
esta: 1
creciente: 1
sofisticación: 1
cuestiones: 1
éticas: 1
privacidad: 1
deben: 1
ser: 1
abordadas: 1
cuidado,: 1
ya: 1
puede: 1
utilizarse: 1
inapropiada: 1
o: 1
fines: 1
perjudiciales.: 1
A: 1
medida: 1
continúa: 1
evolucionando,: 1
espera: 1
tenga: 1
impacto: 1
significativo: 1
amplia: 1
variedad: 1
industrias,: 1
desde: 1
atención: 1
médica: 1
educación: 1
hasta: 1
industria: 1
automotriz: 1
logística.: 1
También: 1
desafíos: 1
importantes: 1
términos: 1
regulación: 1

```
responsabilidad,: 1
lo: 1
hace: 1
ética: 1
sea: 1
tema: 1
crucial: 1
desarrollo: 1
futuro.: 1
En: 1
última: 1
instancia,: 1
tiene: 1
potencial: 1
eficiencia: 1
calidad: 1
vida,: 1
pero: 1
requiere: 1
gestión: 1
cuidadosa: 1
garantizar: 1
utilice: 1
beneficiosa: 1
responsable.: 1
```

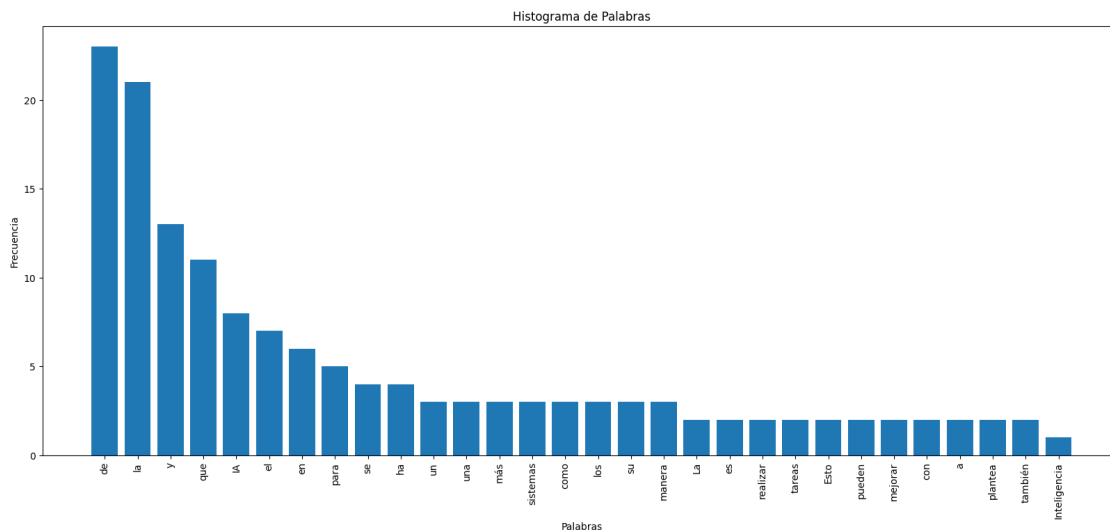
```
[69]: '''
      Se genera un histograma de frecuencia de las palabras identificadas en el texto
      '''
      palabras, frecuencias = zip(*diccionario.items())
      plt.figure(figsize=(20,8))
      plt.bar(palabras, frecuencias)
      plt.xlabel('Palabras')
      plt.ylabel('Frecuencia')
      plt.title('Histograma de Palabras')
      plt.xticks(rotation=90, fontsize=7)
      plt.show()
```



```
[71]: '''
      Se genera un histograma de frecuencia de las primeras 30 palabras con mayor
      frecuencia identificadas en el texto
      '''

      palabras = list(diccionario.keys())[:30]
      frecuencias = [diccionario[palabra] for palabra in palabras]

      plt.figure(figsize=(20,8))
      plt.bar(palabras, frecuencias)
      plt.xlabel('Palabras')
      plt.ylabel('Frecuencia')
      plt.title('Histograma de Palabras')
      plt.xticks(rotation=90, fontsize=10)
      plt.show()
```



```
[72]: '''
      Se genera la nube de palabras de los primeros 30 items del diccionario
      '''

      wordcloud = WordCloud(width=800, height=400).
      ↪generate_from_frequencies(diccionario)
      plt.figure(figsize=(10, 5))
      plt.imshow(wordcloud, interpolation='bilinear')
      plt.axis("off")
      plt.show()
```