Programas en phyton Modulo NLTK

programa 1

```
# -*- coding: utf-8 -*-


'''

El ejercicio consiste en encontrar todas las "palabras" de 3 o 4 letras

- Se entiende por "palabra" CUALQUIER coosa entre espacios

'''


import re


carpeta_nombre="D:\\oswaldo\\FIME ENE-AGO 2022\\PLN\\programas-phyton\\Documentos\\"

archivo_nombre="documento2.txt"


with open(carpeta_nombre+archivo_nombre,"r") as archivo:

        texto=archivo.read()


expresion_regular=re.compile(r"...? ")


resultados_busqueda=expresion_regular.finditer(texto)

for resultado in resultados_busqueda:

        print(resultado.group(0))
```

programa 2

```
import nltk
```

```
import matplotlib.pyplot as plt
carpeta_nombre="D:\\oswaldo\\FIME ENE-AGO 2022\\PLN\\programas-phyton\\Documentos\\"
archivo_nombre="Procesamiento de Lenguaje Natural 1.txt"
with open(carpeta_nombre+archivo_nombre,"r") as archivo:
    texto=archivo.read()
print("------------------------------------------------------------------------")
tokens=nltk.word_tokenize(texto, "spanish")

tokens_conjunto=set(tokens)
palabras_totales=len(tokens)
palabras_diferentes=len(tokens_conjunto)
print(palabras_totales)
print(palabras_diferentes)
texto_nltk=nltk.Text(tokens)
distribucion=nltk.FreqDist(texto_nltk)
print("------------------------------------------------------------------------")
hapaxes=distribucion.hapaxes()
for hapax in hapaxes:
    print(hapax)
from matplotlib import rcParams
rcParams.update({"figure.autolayout": True})
distribucion.plot(cumulative=True)
distribucion.plot(40,cumulative=True)
```

programa 3

```
'''Aqui escribe tu nombre'''
import nltk
print("aqui tambien escribe tu nombre")
carpeta_nombre="F:\\oswaldo\\FIME ENE-AGO 2022\\PLN\\programas-phyton\\Documentos\\"
archivo_nombre="Procesamiento de Lenguaje Natural 1.txt"
with open(carpeta_nombre+archivo_nombre,"r") as archivo:
    texto=archivo.read()
print("------------------------------------------------------------------------")
palabras_funcionales=nltk.corpus.stopwords.words("spanish")
for palabras_funcional in palabras_funcionales:
    '''print(palabras_funcional)'''
print("------------------------------------------------------------------------")
tokens=nltk.word_tokenize(texto,"spanish")
tokens_limpios=[]
for token in tokens:
    if token not in palabras_funcionales:
        tokens_limpios.append(token)
        '''print(tokens_limpios)'''
print(len(tokens))
print(len(tokens_limpios))
texto_limpio_nltk=nltk.Text(tokens_limpios)
distribucion_limpia=nltk.FreqDist(texto_limpio_nltk)
distribucion_limpia.plot(40)
```

programa 4

```
# -*- coding: utf-8 -*-


'''

El ejercicio consiste en encontrar todos los documentos DOC y DOCX de una carpeta usando expresiones regulares.

'''


import os

import re



carpeta_nombre="D:\\oswaldo\\FIME ENE-AGO 2022\\PLN\\programas-phyton\\Documentos\\"



archivos_lista=os.listdir(carpeta_nombre)


expresion_regular=re.compile(r"\.docx?$")


for archivo_nombre in archivos_lista:
        resultado_busqueda=expresion_regular.search(archivo_nombre)
        if resultado_busqueda:
                print(resultado_busqueda.group(0))
                print(archivo_nombre)
```