

Actividad NLP - 1.0 - Creación de Diccionario

Isai Ambrocio - A01625101

Librerías e instalaciones.

```
!pip install wordcloud nltk numpy matplotlib

import matplotlib.pyplot as plt
import nltk
import numpy as np
import re
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from wordcloud import WordCloud

nltk.download("punkt")
nltk.download("stopwords")
```

Funciones para conteo de palabras(`contar_palabras`) y creación de wordcloud(`generar_wordcloud`).

```
def contar_palabras(texto):
    """
    Realiza la tokenización y eliminación de puntuaciones,
    y devuelve las veces que se encontró la palabra
    en el texto.

    :param texto: archivo de texto(.txt).
    :return: frecuencias
    """

    # Tokenización y eliminación de puntuaciones.
    palabras = word_tokenize(texto)
    palabras = [re.sub(r'[^\w\s]', '', palabra) for palabra in palabras]

    # Eliminación de preposiciones y palabras vacías
    stop_words = set(stopwords.words("spanish"))
    palabras = [palabra.lower() for palabra in palabras if
palabra.lower() not in stop_words]

    # Crear un diccionario de frecuencias
    frecuencias = {}
    for palabra in palabras:
        if palabra in frecuencias:
            frecuencias[palabra] += 1
        else:
            frecuencias[palabra] = 1
```

```

    return frecuencias

def generar_wordcloud(frecuencias):
    """
    Toma el diccionario de frecuencias como
    entrada y crea un objeto de nube de palabras.

    :param frecuencias: Diccionario de la función `contar_palabras`
    :return:
    """

    # Configuración de las características para mostrar el wordcloud.
    wordcloud = WordCloud(width = 1024, height = 480, background_color
= "white",
                           max_words = 200, contour_width = 3,
                           contour_color="steelblue", random_state=42)

    wordcloud.generate_from_frequencies(frecuencias = frecuencias)

    plt.figure(figsize = (10, 5))
    plt.imshow(wordcloud, interpolation = "bilinear")
    plt.axis("off")
    plt.show()

```

Abrir archivo de texto y llamar a las funciones.

```

with open("tarea_nlp.txt", "r", encoding="utf-8") as archivo:
    texto = archivo.read()

# Contar palabras
frecuencias = contar_palabras(texto)

# Generar y mostrar WordCloud
generar_wordcloud(frecuencias)

```

agua vida cordero dónde verdadera desatar cielo envío quiere verbo cierto santo volviéndose ser hablar dos negó hijos unigénito espíritu paloma plenitud tomamos decima nazaret bautizando lleno

habito quita decir pedro manifestado seno así padre jonás cristo pues simón dice dado mirando hallado engendrados discípulos engendrados alumbra hombre despusés

hechas dada ved gloria sino felipe respondió moisés hermano sacerdotes después andrés

si verdad elias preguntaron decia bautiza moises tinieblas israel sigue me conocéis

medio gracia venia confesó siguiente vimo potestad mas aquí dices

principio andaba señor llamado creyesen vinieron voluntad jordan jamas recibieron oído calzado isaías diciendo

debajo varón hora bautizo escribio jordan jamas recibieron oído calzado isaías diciendo

cosas llamaba conocía testimonio bautizo escribio jordan jamas recibieron oído calzado isaías diciendo

profeta