**PROGRAMA PROFESIONAL**

Ciencias de la Computación

**TÍTULO DEL TRABAJO**

Informe de implementación de algoritmo de Porter

**CURSO**

Base de datos II

**ALUMNOS**

* Angel Josue Loayza Huarachi

**SEMESTRE:** V

**AÑO:** 2022

“El alumno declara haber realizado el presente trabajo de acuerdo a las normas de la Universidad Católica San Pablo

Contenido

[1. Herramientas usadas 3](#_Toc100617676)

[2. Lista de archivos creados 3](#_Toc100617677)

[3. Implementación Algoritmo de Porter 3](#_Toc100617678)

[3.1. Archivo “texto.txt” 3](#_Toc100617679)

[3.2. Archivo “texto2.txt” 4](#_Toc100617680)

[3.3. Archivo “test.py” 4](#_Toc100617681)

[3.4. Archivo “Untitled.ipynb” 4](#_Toc100617682)

[3.4.1. Resultado al usar el archivo “texto.txt” 5](#_Toc100617683)

[3.4.2. Resultado al usar el archivo “texto2.txt” 6](#_Toc100617684)

[4. Enlace a GitHub 6](#_Toc100617685)

# Herramientas usadas

* Natural Language Toolkit (NLTK)
* Jupyter notebook

# Lista de archivos creados

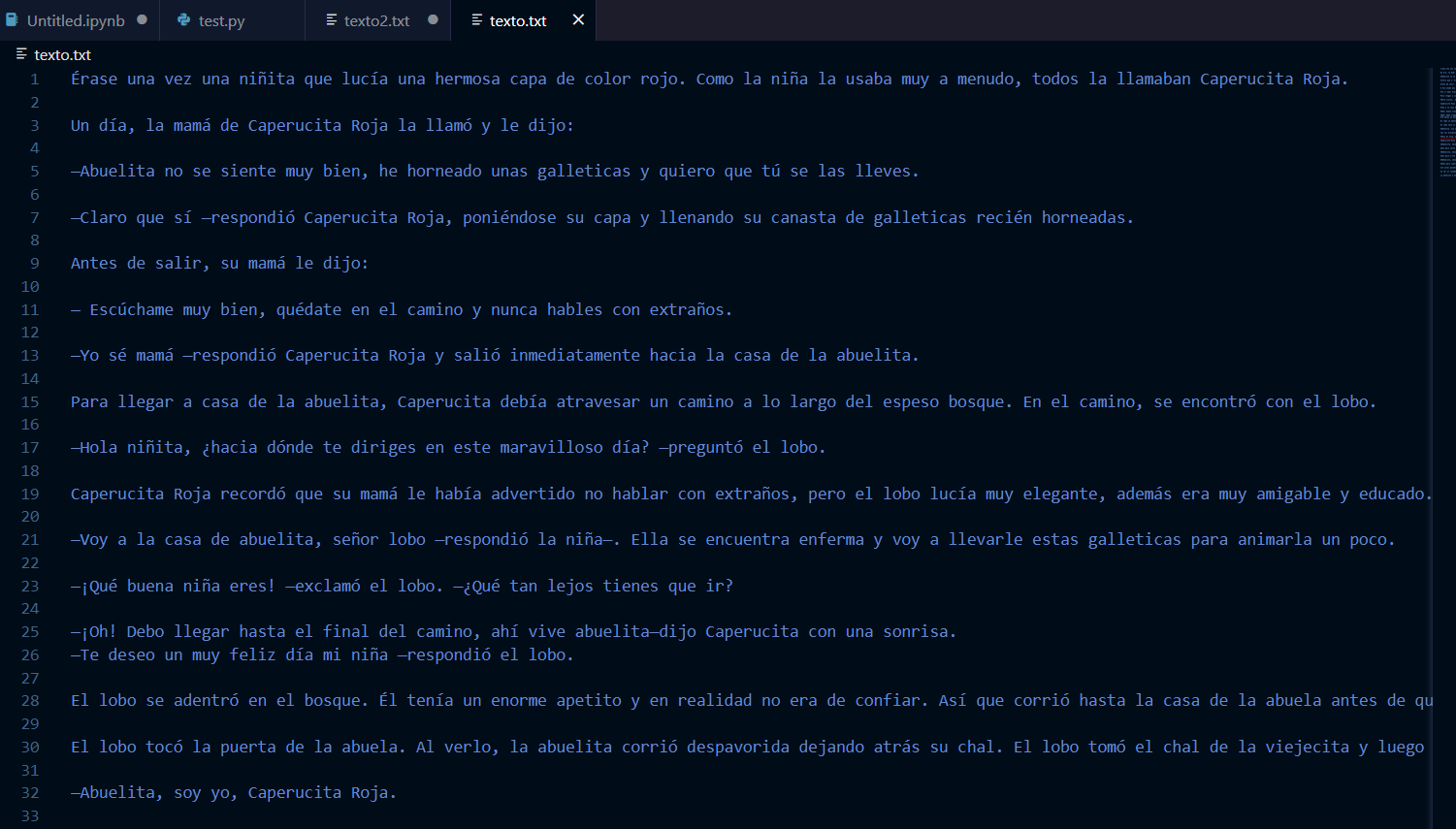
* texto.txt
  + Contiene el cuento “La Caperucita Roja” extraído del Moodle
* texto2.txt
  + Contiene el cuento “La Caperucita Roja” en ingles extraído del siguiente link:

<http://eidioma.blogspot.com/2011/01/caperucita-roja-en-ingles.html>

* test.py
  + Contiene un archivo de prueba de lectura y escritura de los 2 txt’s que se usó en el archivo main
* Untitled.ipynb
  + Contiene el código principal donde se importa las librerías de Porter y realiza la lectura, escritura y uso de dicha librería al documento de texto

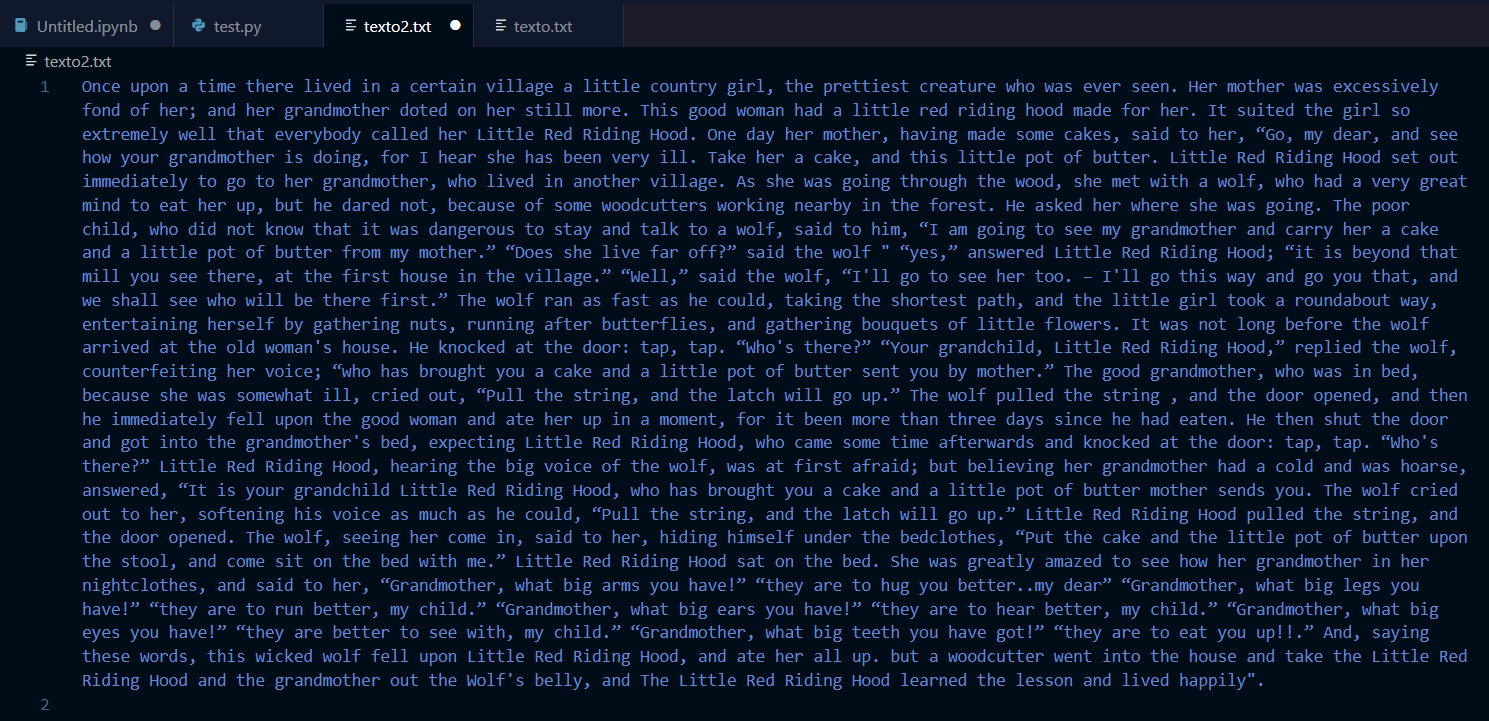
# Implementación Algoritmo de Porter

## Archivo “texto.txt”

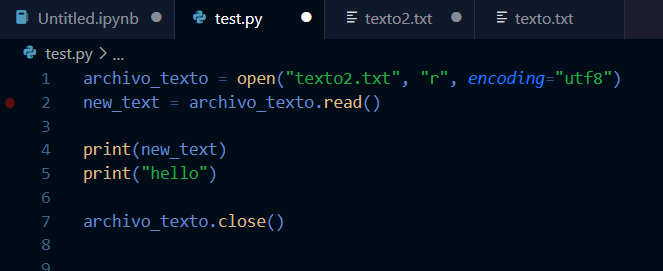


**Dato: Solo se mostro una parte del archivo, en el enlace a GitHub se puede presenciar todo el cuento completo**

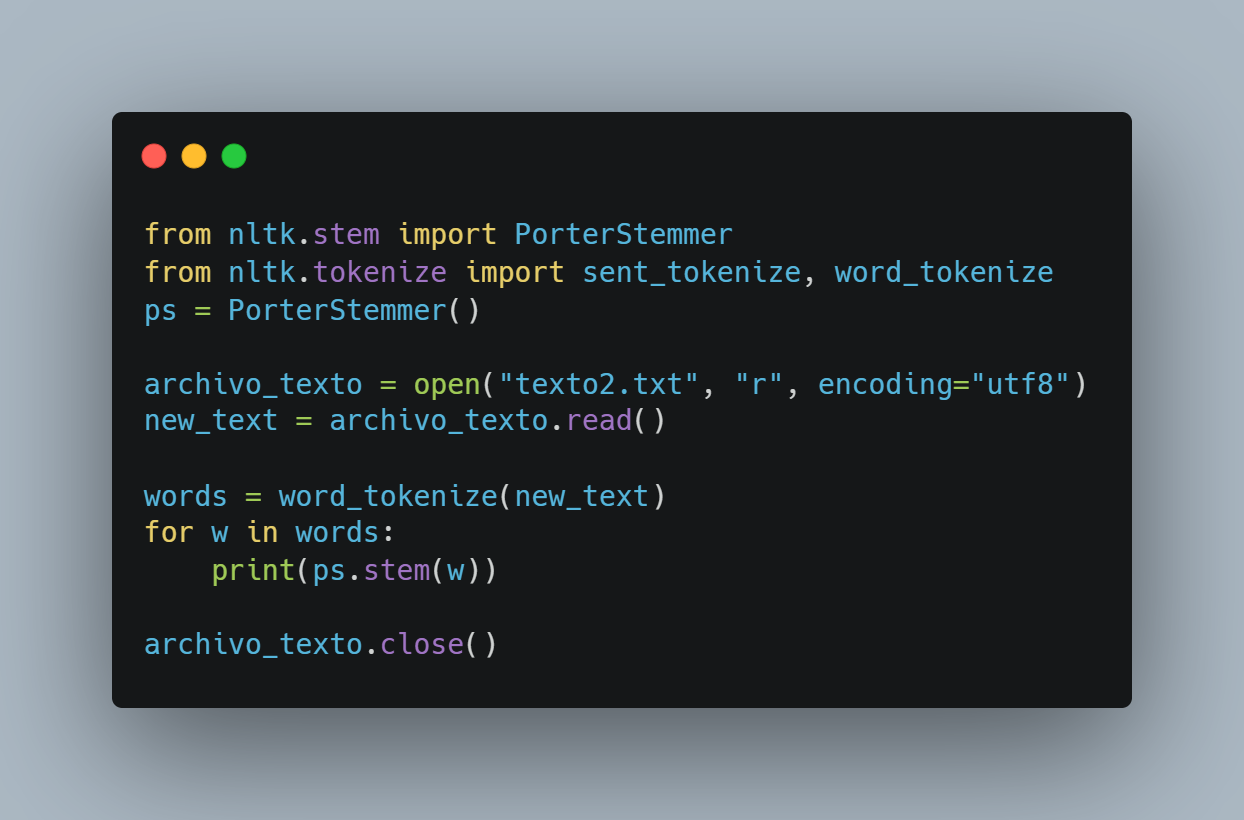
## Archivo “texto2.txt”



## Archivo “test.py”



## Archivo “Untitled.ipynb”



# --------------------------- Codigo --------------------------------------------

from nltk.stem import PorterStemmer

from nltk.tokenize import sent\_tokenize, word\_tokenize

ps = PorterStemmer()

archivo\_texto = open("texto2.txt", "r", encoding="utf8")

new\_text = archivo\_texto.read()

words = word\_tokenize(new\_text)

for w in words:

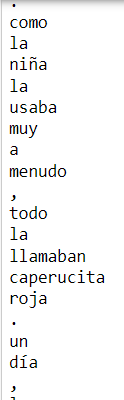
print(ps.stem(w))

archivo\_texto.close()

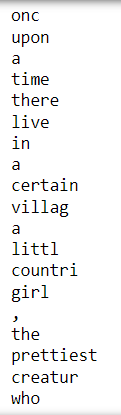
# ------------------------------------------------------------------------------------

**Dato: Los resultados mostrados a continuación son un extracto de la respuesta completa. Los archivos con los resultados completos estarán en el enlace a GitHub.**

### Resultado al usar el archivo “texto.txt”



### Resultado al usar el archivo “texto2.txt”



# Enlace a GitHub

<https://github.com/angel452/BaseDeDatosII>