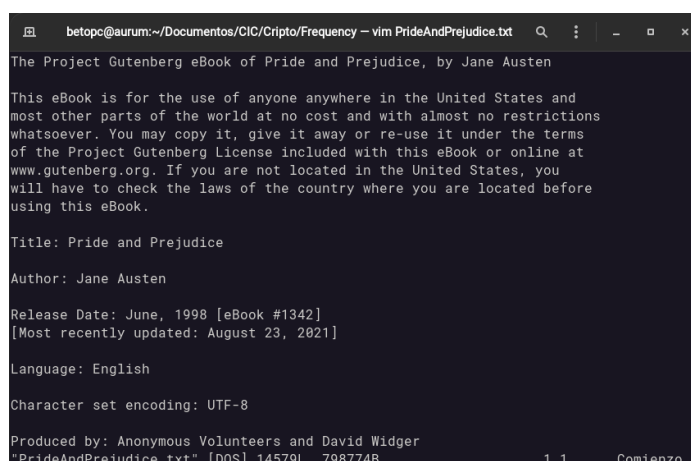


Introducción a la Criptografía (B22)

Tarea 2: Análisis de frecuencia de aparición de símbolos

Palacios Cabrera Alberto

El primer paso para el funcionamiento del programa es obtener un texto para analizar en formato *txt* compuesto principalmente por símbolos pertenecientes al alfabeto inglés (26 letras en el alfabeto). En este caso se utilizaron textos de libros obtenidos por medio de “*Project Gutenberg*” (repositorio con una basta cantidad de libros cuyo copyright ha expirado y por lo cual son de dominio público). En la Figura 1 se muestra el texto del libro “*Orgullo y prejuicio*” con el cual se realizaron las pruebas del programa.



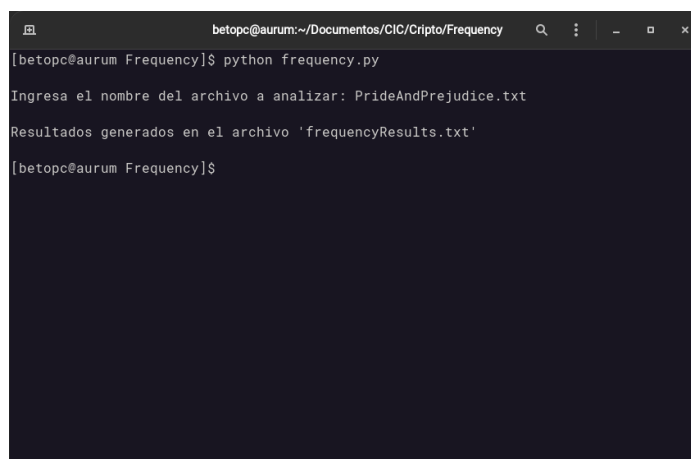
```
betopc@aurum:~/Documentos/CIC/Cripto/Frequency - vim PrideAndPrejudice.txt
The Project Gutenberg eBook of Pride and Prejudice, by Jane Austen

This eBook is for the use of anyone anywhere in the United States and
most other parts of the world at no cost and with almost no restrictions
whatsoever. You may copy it, give it away or re-use it under the terms
of the Project Gutenberg License included with this eBook or online at
www.gutenberg.org. If you are not located in the United States, you
will have to check the laws of the country where you are located before
using this eBook.

Title: Pride and Prejudice
Author: Jane Austen
Release Date: June, 1998 [eBook #1342]
[Most recently updated: August 23, 2021]
Language: English
Character set encoding: UTF-8
Produced by: Anonymous Volunteers and David Widger
"PrideAndPrejudice.txt" [DOS] 14579L, 798774B          1,1      Comienzo
```

Figura 1 Archivo txt “PrideAndPrejudice.txt”

En la ejecución del programa se solicita al usuario que ingrese el nombre del archivo a analizar. El nombre del archivo debe especificar la ruta de este en caso de ubicarse fuera de la carpeta donde se encuentra el programa. Una vez que el programa termina el conteo de símbolos y los cálculos pertinentes, indica al usuario que los resultados se encuentran en el archivo “*frequencyResults.txt*” El programa en ejecución se muestra en la Figura 2.



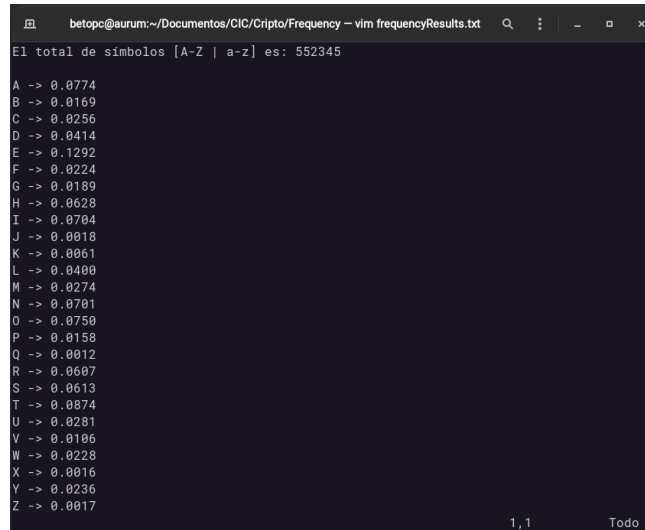
```
betopc@aurum:~/Documentos/CIC/Cripto/Frequency
[betopc@aurum Frequency]$ python frequency.py

Ingresa el nombre del archivo a analizar: PrideAndPrejudice.txt
Resultados generados en el archivo 'frequencyResults.txt'

[betopc@aurum Frequency]$
```

Figura 2 Ejecución de frequency.py

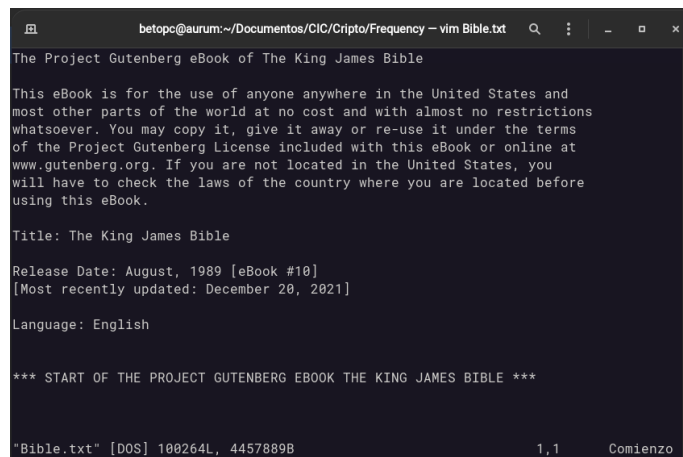
Por último, los resultados del análisis de frecuencia de aparición de símbolos de muestra en el archivo de texto ***“frequencyResults.txt”***. En este archivo se muestra al usuario la cantidad de símbolos que fueron contados (únicamente considerando aquellos que se encuentran en el alfabeto inglés) y también un listado de pares ordenados en el cual se indica el símbolo encontrado y su frecuencia de aparición en el texto analizado. Este archivo se muestra en la Figura 3.



```
betopc@aurum:~/Documentos/CIC/Cripto/Frequency - vim frequencyResults.txt
El total de símbolos [A-Z | a-z] es: 552345
A -> 0.0774
B -> 0.0169
C -> 0.0256
D -> 0.0414
E -> 0.1292
F -> 0.0224
G -> 0.0189
H -> 0.0628
I -> 0.0704
J -> 0.0018
K -> 0.0061
L -> 0.0400
M -> 0.0274
N -> 0.0701
O -> 0.0750
P -> 0.0158
Q -> 0.0012
R -> 0.0607
S -> 0.0613
T -> 0.0874
U -> 0.0281
V -> 0.0106
W -> 0.0228
X -> 0.0016
Y -> 0.0236
Z -> 0.0017
1,1 Todo
```

Figura 3 Archivo *frequencyResults.txt*

A continuación, en la Figura 4, Figura 5 y Figura 6 se muestra la prueba realizada con el texto ***“The King James Bible”***.



```
betopc@aurum:~/Documentos/CIC/Cripto/Frequency - vim Bible.txt
The Project Gutenberg eBook of The King James Bible

This eBook is for the use of anyone anywhere in the United States and
most other parts of the world at no cost and with almost no restrictions
whatsoever. You may copy it, give it away or re-use it under the terms
of the Project Gutenberg License included with this eBook or online at
www.gutenberg.org. If you are not located in the United States, you
will have to check the laws of the country where you are located before
using this eBook.

Title: The King James Bible

Release Date: August, 1989 [eBook #10]
[Most recently updated: December 20, 2021]

Language: English

*** START OF THE PROJECT GUTENBERG EBOOK THE KING JAMES BIBLE ***

"Bible.txt" [DOS] 100264L, 4457889B
1,1 Comienzo
```

Figura 4 Archivo txt *“Bible.txt”*

```
betopc@aurum:~/Documentos/CIC/Cripto/Frequency
[betopc@aurum Frequency]$ python frequency.py
Ingresa el nombre del archivo a analizar: Bible.txt
Resultados generados en el archivo 'frequencyResults.txt'
[betopc@aurum Frequency]$
```

Figura 5 Prueba de frequency.py con "Bible.txt"

```
betopc@aurum:~/Documentos/CIC/Cripto/Frequency - vim frequencyResults.txt
El total de símbolos [A-Z | a-z] es: 3241189
A -> 0.0851
B -> 0.0151
C -> 0.0170
D -> 0.0488
E -> 0.1273
F -> 0.0258
G -> 0.0171
H -> 0.0873
I -> 0.0599
J -> 0.0027
K -> 0.0069
L -> 0.0401
M -> 0.0247
N -> 0.0695
O -> 0.0751
P -> 0.0134
Q -> 0.0003
R -> 0.0526
S -> 0.0507
T -> 0.0981
U -> 0.0258
V -> 0.0094
W -> 0.0202
X -> 0.0005
Y -> 0.0181
Z -> 0.0009
1,1 Todo
```

Figura 6 Resultados de prueba con "Bible.txt"