

Tarea 1: Teoría de la Información

Resumen:

Se realizará un script en Python para concretar la teoría revisada acerca de la Información. Una tarea anterior será utilizada para obtener los códigos de Huffman de un determinado texto en un archivo .txt. La prueba fue realizada en un archivo de 10333 caracteres.

Introducción:

Se reportará los resultados de un archivo .txt de 10 000 caracteres. Estos resultados son los códigos de Huffman para el alfabetos del mensaje, y se obtendrán diversos parámetros de interés para la criptografía.

Descripción del problema:

D de un mensaje dado cual sea el tamaño, obtener su codificación de Huffman para el alfabeto de dicho mensaje, se obtendrán otros parametros de interés como el rango absoluto y real, codificación óptima, redundancia, información que aporta cada caracter, y la entropía del texto de entrada.

Descripción del programa desarrollado en Python:

El programa cuenta con la obtención de los códigos de Huffman, óptimos para el alfabeto estudiado del mensaje. Muestra además, la longitud del mensaje, la función de distribución de probabilidad, la codificación de cada caracter, además de los parámetros solicitados, por último imprime la cantidad de bits que aporta cada caracter dada su probabilidad.

El script hace un llamado a “LongText.txt” para comenzar a estudiarlo.

Resultados:

Longitud del mensaje: 10333

Longitud de alfabeto: 67

Histograma y codificación:

T : 0.0021291009387399596 : 110100101

e : 0.08864802090390013 : 000

x : 0.002612987515726314 : 00110001

t : 0.0681312300396787 : 1010
: 0.15445659537404433 : 111
c : 0.033872060389044806 : 10011
l : 0.0322268460272912 : 10010
a : 0.06861511661666506 : 1011
s : 0.06126004064647247 : 0110
i : 0.06542146520855512 : 1000
f : 0.021291009387399595 : 110111
o : 0.06958288977063776 : 1100
n : 0.05825994386915707 : 0100
h : 0.029129971934578534 : 01010
m : 0.024678215426304075 : 00111
p : 0.02080712281041324 : 110110
r : 0.04732410722926546 : 0010
k : 0.0020323236233426886 : 011111101
N : 0.0003871092615890835 : 00110100101
u : 0.018871576502467822 : 110101
L : 0.001645214361753605 : 001101110
g : 0.014516597309590632 : 011100
P : 0.000774218523178167 : 0011010110
.: 0.007548630600987129 : 0111010
I : 0.0021291009387399596 : 110100100
y : 0.01538759314816607 : 011110
d : 0.029613858511564888 : 01011
, : 0.00890351301654892 : 1101000
v : 0.0060001935546307944 : 0011001
w : 0.00783896254717894 : 0111011
b : 0.008032517177973483 : 0111110
q : 0.0021291009387399596 : 011111111
Salto : 9.677731539727088e-05 : 0011011110010
- : 0.0009677731539727088 : 0011011111
W : 0.0002903319461918126 : 011111110011
' : 0.0003871092615890835 : 00110100100
M : 0.0010645504693699798 : 1101001100
1 : 0.0004838865769863544 : 01111111000
z : 0.0008709958385754379 : 0011010111
(: 0.001645214361753605 : 001101101
) : 0.001645214361753605 : 001101100
A : 0.0020323236233426886 : 011111100
: : 0.0005806638923836253 : 0011000000
" : 0.0013548824155617924 : 001100001
D : 0.0010645504693699798 : 0111111101

C : 0.002516210200329043 : 110100111
H : 0.0002903319461918126 : 011111110010
S : 0.000774218523178167 : 0011010001
9 : 0.00019355463079454176 : 001101111000
4 : 0.0002903319461918126 : 110100110101
8 : 0.00019355463079454176 : 001101000001
V : 9.677731539727088e-05 : 00110111100111
E : 0.0006774412077808962 : 0011000001
F : 0.001548437046356334 : 001101010
B : 0.00019355463079454176 : 001101000000
Z : 0.0002903319461918126 : 110100110100
G : 0.0004838865769863544 : 00110111101
U : 0.0002903319461918126 : 110100110111
/ : 9.677731539727088e-05 : 00110111100110
j : 9.677731539727088e-05 : 0011010011001
3 : 0.0002903319461918126 : 110100110110
0 : 0.0003871092615890835 : 00110100111
- : 0.00019355463079454176 : 001101000011
2 : 9.677731539727088e-05 : 0011010011000
% : 0.00019355463079454176 : 001101000010
5 : 9.677731539727088e-05 : 0011010011011
+ : 9.677731539727088e-05 : 0011010011010

Entropia: 4.3674085136667316 bits/letter

Rango Absoluto: 0.06518520169651838 bits/letter

Redundancia: 4.302223311970213

Promedio de bits(Huffman Codes): 5.003483983354298

Información por caracter:

T 8.875539936390163
e 3.4957677669305185
x 8.580084052863993
t 3.875539936390165
2.694726618805116
c 4.883760443195134
l 4.9555931879562
a 3.8653297377879463
s 4.028909865599121
i 3.9340921187452778
f 5.553611841502803
o 3.845123594588165
n 4.10135187826776
h 5.10135187826776
m 5.340618118168605

p 5.586778705438002
r 4.401280900075228
k 8.942654132248702
N 11.334971555027463
u 5.727641241277852
L 9.247508713777123
g 6.106152864531581
P 10.334971555027463
. 7.049569336165214
I 8.875539936390163
y 6.022088599743107
d 5.07758371233481
, 6.81140959897045
v 7.380775244640587
w 6.995121552142838
b 6.959932123680538
q 8.875539936390163
Salto 13.334971555027462
- 10.0130434601401
W 11.750009054306307
' 11.334971555027463
M 9.875539936390165
1 11.0130434601401
z 10.16504655358515
(9.247508713777123
) 9.247508713777123
A 8.942654132248702
: 10.750009054306306
" 9.527616632969858
D 9.875539936390165
C 8.63453183688637
H 11.750009054306307
S 10.334971555027463
9 12.334971555027462
4 11.750009054306307
8 12.334971555027462
V 13.334971555027462
E 10.52761663296986
F 9.334971555027463
B 12.334971555027462
Z 11.750009054306307
G 11.0130434601401

U 11.750009054306307
/ 13.334971555027462
j 13.334971555027462
3 11.750009054306307
0 11.334971555027463
– 12.334971555027462
2 13.334971555027462
% 12.334971555027462
5 13.334971555027462
+ 13.334971555027462

Entropía: 4.3674085136667316 AvBits: 5.003483983354298 Entropía + 1:
5.3674085136667316

Conclusiones:

Estos parámetros estudiados se deben abordar de manera práctica y teórica para realizar en futuros trabajos un sencillo sistema de encriptación, midiendo estos parámetros se puede determinar que tan bien hecho está algún criptosistema. Se aprendieron los conceptos y fundamentos más básicos a analizar en alguna encriptación.

Referencias:

Charles K. **Chui** Qingtang **Jiang**, *Applied Mathematics Data Compression, Spectral Methods, Fourier Analysis, Wavelets, and Applications*.