clc; close all; clear all;

%ABRIR ARCHIVO

archivo=fopen('/principitoe.txt','r');

cadena = fscanf(archivo,'%c'); %guardo archivo en un vector

numcaracteres=length(cadena); %numero de caracteres del archivo

fclose(archivo);

%CIERRE DE ARCHIVO

%\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

%\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

% ANALISIS POR CARACTER

%vector con las letras del abecedario incluyendo espacio

ABC=['a','b','c','d','e','f','g','h','i','j','k','l','m','n','ñ','o','p','q','r','s','t','u','v','w','x','y','z',' '];

%conteo de repeticiones por cada caracter

ABC0=zeros(1,length(ABC)); %vector para el conteo de repeticiones de cada letra

cont=1;

for x=1:length(ABC);

rep=0; %variable donde se guarda el no. de repeticiones de 'x' caracter

for y=1:length(cadena);

if cadena(y)==ABC(x);

rep=rep+1;

end

end

ABC0(cont)=rep; %guardo el no. de repeticiones en el vector de conteo

cont=cont+1;

end

p1=zeros(1,length(ABC0));

I1=zeros(1,length(ABC0));

for x=1:length(ABC0);

p=(ABC0(x)./sum(ABC0));%probabilidad=cantidad de repeticiones/no. de caracteres

p1(x)=p;

I=log2(1/p);% cantidad de informacion por caracter

if I==Inf %cuando la P(x)=0 la I(x)=inf

I=0; %se hace esta igualacion para que no se invalide el calculo

end %de la entropia por los inf

I1(x)=I;

end

%calculo de la entropia como fuente sin memoria

h=zeros(1,length(p1));

for x=1:length(p1);

h(x)=p1(x)\*I1(x);

end

H=sum(h);

fprintf('\nESPAÑOL: ENTROPIA COMO FUENTE SIN MEMORIA= %f \n',H);

%\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

%\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

% ANALISIS PARA PAREJAS

i=1; j=1;

cont2=zeros(length(ABC),length(ABC)); %matriz de ceros que llevara el conteo de repeticiones

for x=1:length(ABC);

for y=1:length(ABC); %generamos una matriz para hacer las combinaciones

cont=0; %lleva el conteo de repeticiones por parejas

for z=1:(length(cadena)-1);

if cadena(z)==ABC(x) && cadena(z+1)==ABC(y); %barre la cadena

cont=cont+1;

end

end

cont2(i,j)=cont;

j=j+1; %barre columnas de la matriz

end

j=1;

i=i+1; %barre filas de la matriz

end

pares=zeros(1,(length(ABC)^2));

i=1;

for x=1:length(ABC); %barremos la matriz con los valores de repeticion

for y=1:length(ABC);

if cont2(x,y)~=0; %solo se imprimen los valores diferenes de 0

txt=[ ABC(x),ABC(y) ' repeticiones: ' int2str(cont2(x,y))];

end

pares(1,i)=cont2(x,y); % se guarda un vector con el valor de las repeticiones

i=i+1;

end

end

p2=zeros(1,length(pares));

I2=zeros(1,length(pares));

%calculo de probabilidades para pares

for x=1:length(pares);

p=(pares(x)./sum(pares));%probabilidad=cantidad de repeticiones/no. de caracteres

p2(x)=p;

I=log2(1/p);% cantidad de informacion por caracter

if I==Inf %cuando la P(x)=0 la I(x)=inf

I=0; %se hace esta igualacion para que no se invalide el calculo

end %de la entropia por los inf

I2(x)=I;

end

maximos=pares;

max=sort(maximos,'descend');

for x=1:10;

max10(x)=max(x);

end

%\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

%\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

% ANALISIS PARA TERCIAS

i=1; j=1; k=1;

cont3=zeros(length(ABC),length(ABC),length(ABC));

for x=1:length(ABC);

for y=1:length(ABC);

for z=1:length(ABC);

cont=0;

for a=1:length(cadena)-2; %barre cadena

if cadena(a)==ABC(x) && cadena(a+1)==ABC(y) && cadena(a+2)==ABC(z);

cont=cont+1;

end

end

cont3(i,j,k)=cont;

k=k+1;

end

k=1;

j=j+1;

end

j=1;

i=i+1;

end

i=1;

for x=1:length(ABC);

for y=1:length(ABC);

for z=1:length(ABC);

if cont3(x,y,z)~=0;

txt=[ ABC(x),ABC(y), ABC(z) ' repeticiones: ' int2str(cont3(x,y,z))];

end

tercias(1,i)=cont3(x,y,z);

i=i+1;

end

end

end

p3=zeros(1,length(tercias));

I3=zeros(1,length(tercias));

%calculo de probabilidades para tercias

for x=1:length(tercias);

p=(tercias(x)./sum(tercias));%probabilidad=cantidad de repeticiones/no. de caracteres

p3(x)=p;

I=log2(1/p);% cantidad de informacion por caracter

if I==Inf %cuando la P(x)=0 la I(x)=inf

I=0; %se hace esta igualacion para que no se invalide el calculo

end %de la entropia por los inf

I3(x)=I;

end

maximos2=tercias;

max2=sort(maximos2,'descend');

for x=1:10;

max102(x)=max2(x);

end

%\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

%\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

% MARKOV DE 1er ORDEN

pc1=zeros(length(ABC),length(ABC));

for x=1:length(ABC);

for y=1:length(ABC);

if ABC0(y)==0;

pc1(x,y)=0;

else

pc1(x,y)=cont2(x,y)/ABC0(y); %suma todo el vector x

end

end

end

marcov1=0;

for x=1:length(ABC);

for y=1:length(ABC);

if pc1(x,y)~=0;

marcov1=marcov1 + ( p1(y)\*pc1(x,y)\*(log2(1/pc1(x,y))) );

end

end

end

fprintf('\nESPAÑOL: ENTROPIA COMO FUENTE CON MEMORIA MARKOV 1°= %f \n',marcov1);

%\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

%\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

% MARKOV DE 2do ORDEN

pc2=zeros(length(ABC),length(ABC),length(ABC));

for x=1:length(ABC);

for y=1:length(ABC);

for z=1:length(ABC);

if cont3(x,y,z)==0;

pc2(x,y,z)=0;

else

pc2(x,y,z)=cont3(x,y,z)/cont2(y,z); %suma todo el vector x,y,:

end

end

end

end

marcov2=0;

for x=1:length(ABC);

for y=1:length(ABC);

for z=1:length(ABC);

if pc2(x,y,z)~=0;

marcov2=marcov2+(p1(z)\*pc1(y,z)\*pc2(x,y,z)\*(log2(1/pc2(x,y,z))));

end

end

end

end

fprintf('\nESPAÑOL: ENTROPIA COMO FUENTE CON MEMORIA MARKOV 2°= %f \n\n',marcov2);

figure(1)

hs=[H,marcov1,marcov2];

bar(hs,'m')