

TEORÍA DE LA INFORMACIÓN Y LA CODIFICACIÓN

Hollman Leonardo Pinto Zapata, Ivanna Torres Perez

Abstract—En el siguiente documento se muestra una breve explicación de la manera en que se codifica la información en binario que luego se va a transmitir, sus probabilidades de ocurrencia, entropía y otros parámetros importantes que se deben tener en cuenta al momento de codificar y enviar un mensaje.

Index Terms—Entropía, Codificación, Incertidumbre.

I. INTRODUCCIÓN

La necesidad de comunicación implícita dentro del ser humano y por ende la sociedad ha llevado al desarrollo de sistemas y protocolos de comunicación que cumplen diferentes tareas y específicos para ciertos mensajes, para hacer mas eficiente la comunicación se han desarrollado a la par de la codificación binaria métodos como la codificación de canal y de fuente, la codificación Huffman o de código compacto. en este sentido se puede hablar de la teoría de la información. La Teoría de la Información de Claude E. Shannon, es sin duda uno de los avances científicos más importantes del siglo XX. El principal objetivo de esta teoría es el de proporcionar una definición rigurosa de la noción de información que permita cuantificarla. Fue desarrollada con el objetivo de encontrar límites fundamentales en las operaciones de procesamiento de señales tales como compresión de datos, almacenamiento y comunicación. [1]

II. PROCEDIMIENTO

Se desarrolla una interfaz grafica en Microsoft Visual Basic y Excel que permita al usuario ingresar una cadena de caracteres con una cantidad de mínimo 40 con los cuales se desean obtener datos valiosos para su posterior análisis, como lo son: la cantidad de información y la entropía. La interfaz Grafica se observa en la Fig.1

Hollman Pinto pertenece a la Escuela de Ingeniería Electrónica, Sogamoso, Colombia, e-mail: hollman.pinto@uptc.edu.co.

Ivanna Torres pertenece a la Escuela de Ingeniería Electrónica, Sogamoso, Colombia, e-mail: ivanna.torres@uptc.edu.co.

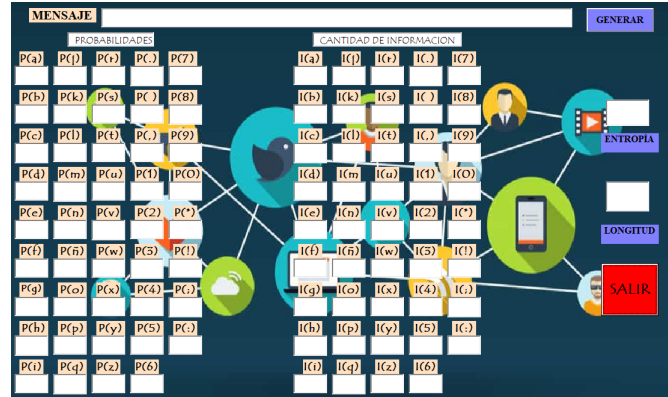


Fig. 1. Interfaz Grafica diseñada en Visual Basic.

Ahora, los datos obtenidos no solamente se muestran en la interfaz grafica, sino que también se guardan en celdas numéricas de una hoja de excel como se observa en la Fig.2 , para luego ser exportados a MATLAB. usando el comando «xlsread» como se observa en la Fig.3

CARGAR		APLICACIÓN	
MENSAGE INGRESADO		1234567890123456789012345678901234567890	
PROBABILIDAD	CANTIDAD DE INFORMACIÓN	ENTROPÍA	LONGITUD
P(a) 0	I(a) 0	3,322	40
P(b) 0	I(b) 0		
P(c) 0	I(c) 0		
P(d) 0	I(d) 0		
P(e) 0	I(e) 0		
P(f) 0	I(f) 0		
P(g) 0	I(g) 0		
P(h) 0	I(h) 0		
P(i) 0	I(i) 0		
P(j) 0	I(j) 0		
P(k) 0	I(k) 0		
P(l) 0	I(l) 0		
P(m) 0	I(m) 0		
P(n) 0	I(n) 0		
P(o) 0	I(o) 0		
P(p) 0	I(p) 0		
P(q) 0	I(q) 0		
P(r) 0	I(r) 0		
P(s) 0	I(s) 0		
P(t) 0	I(t) 0		
P(u) 0	I(u) 0		
P(v) 0	I(v) 0		
P(w) 0	I(w) 0		
P(x) 0	I(x) 0		
P(y) 0	I(y) 0		
P(z) 0	I(z) 0		

Fig. 2. Datos almacenados en formato xls obtenidos de la interfaz grafica.

```
>> num = xlsread('LAB_1.xlsm')
```

Fig. 3. Comando para exportar a MATLAB.

Dos datos importantes se observan en la Fig.2 los cuales son. Entropía y la cantidad de información estos se utilizan para la codificación del canal de transmisión y la manera de calcularlos es la siguiente:

$$H = \sum_{j=1}^n P_j \log_2 \left(\frac{1}{P_j} \right) \quad (1)$$

La entropía se obtiene mediante la Ecuación.1

La cantidad de información se obtiene mediante la Ecuación.2

$$I(s_i) = \log_2\left(\frac{1}{p_i}\right) \quad (2)$$

III. RESULTADOS

3.1 PROBABILIDADES IGUALES:

Para verificar la veracidad de las ecuaciones anteriormente mencionadas y el correcto funcionamiento de la interfaz grafica, se realiza una prueba ingresando la siguiente cadena de caracteres «123456789123456789123456789123456789», para lo cual se obtienen los resultados observados en la Fig.4



Fig. 4. Resultados obtenidos para la cadena de caracteres ingresada.

las probabilidades para cada carácter son iguales ya que para este ejercicio en especial la misma cadena 123456789 se repite 5 veces con lo cual se tienen 45 caracteres, cumpliendo el criterio de mínimo 40

Caracter	Probabilidad
1	5/45
2	5/45
3	5/45
4	5/45
5	5/45
6	5/45
7	5/45
8	5/45
9	5/45

TABLE I

PROBABILIDADES DE OCURRENCIA DE LOS CARACTERES.

al aplicar la ecuación 2 utilizando las probabilidades obtenidas en la Tabla.I se obtiene una cantidad de información de 3.17 bits/Símbolo esto para todos los caracteres debido a que tienen la misma probabilidad de ocurrencia, si esto no fuera así, los números con menos probabilidad de ocurrencia aportarían una mayor cantidad de información con respecto a los que tienen una mayor probabilidad.

aplicando la ecuación. 1 sucede lo mismo que en el caso anterior esto debido a la misma razón de igualdad en las

probabilidades por lo cual al encontrar la entropía va a dar igual a 3.17

3.2 PROBABILIDADES AL AZAR:

Finalmente como lo que se desea verificar es que estas ecuaciones, la teoría y la interfaz grafica funcionen para cualquier cadena de caracteres mayor a 40, entonces se procede a realizar una nueva cadena de prueba con números del 0 al 9 pero totalmente al azar.

Ingresando la siguiente cadena: «157986351294358920483156789942301568435842657» se obtienen los siguientes valores de entropía y cantidad de información.



Fig. 5. Resultados obtenidos para la segunda cadena de caracteres.

Al aplicar la ecuación 2 se obtienen los datos obtenidos en la Tabla.

EVENTO	PROBABILIDAD	CANTIDAD DE INFORMACIÓN
0	2/45	4.492
1	4/45	3.492
2	4/45	3.492
3	5/45	3.169
4	5/45	3.169
5	7/45	2.684
6	4/45	3.492
7	3/45	3.907
8	6/45	2.907
9	5/45	3.169

TABLE II

EVENTOS, PROBABILIDAD Y CANTIDAD DE INFORMACIÓN PARA LA SEGUNDA CADENA DE CARACTERES.

Finalmente para calcular la entropía se utiliza la ecuación.1 de la siguiente manera:

$$\frac{2}{45} \log_2\left(\frac{45}{2}\right) + \frac{4}{45} \log_2\left(\frac{45}{4}\right) + \frac{4}{45} \log_2\left(\frac{45}{4}\right) + \frac{5}{45} \log_2\left(\frac{45}{5}\right) + \frac{5}{45} \log_2\left(\frac{45}{5}\right) + \frac{7}{45} \log_2\left(\frac{45}{7}\right) + \frac{4}{45} \log_2\left(\frac{45}{4}\right) + \frac{3}{45} \log_2\left(\frac{45}{3}\right) + \frac{6}{45} \log_2\left(\frac{45}{6}\right) + \frac{5}{45} \log_2\left(\frac{45}{5}\right) = 3.253$$

IV. CONCLUSIONES

- Como se observa en las tablas I II la información que va a suministrar la fuentes es inversa a la probabilidad

de ocurrencia del evento, es decir si el evento se repite muchas veces ya no va a transmitir información útil, por lo cual se desarrollan canales y receptores que sean capaces de identificar estas situaciones y realizar algún tipo de tarea para optimizar el uso de los recursos disponibles.

- El uso de herramientas tecnologicas como la interfza grafica desarrollada en Excel o el software MATLAB, ayuda a obtener mas rapidamente los valores deseados sin necesidad de aplicar cada ves que se cambia la cadena las ecuaciones a mano, sino que se programan en la interfaz grafica estas ecuaciones y el procesador las evalua mejorando el rendimiento del diseñador.

REFERENCES

- [1] G. L. Castro Ronquillo *et al.*, "Transmisión digital por microondas," B.S. thesis, Espol, 2017.