TEORÍA DE LA INFORMACIÓN Y CODIFICACIÓN

|  |
| --- |
| Jose Daniel Martínez, Fabian Orlando Niño, Steven Horacio Corredor. |
| Sogamoso, Boyacá |
| Universidad Pedagógica y Tecnológica de Colombia |
| Escuela de Ingeniería Electrónica – Seccional Sogamoso  [Josedaniel.martinez@uptc.edu.co](mailto:Josedaniel.martinez@uptc.edu.co), [Fabian.nino@uptc.edu.co](mailto:Fabian.nino@uptc.edu.co), Steven.corredor@uptc.edu.co. |

*Resumen*- En este informe se presenta un programa realizado en Excel con la herramienta de Visual Basic para obtener la entropía de un mensaje de al menos 40 símbolos, también se obtendrá la probabilidad de los caracteres y la cantidad de información, los datos se mostraran en una tabla realizada en Excel.

*Palabras Clave*- Entropía, información, Visual Basic.

Introducción

El propósito de un sistema de comunicación es la transmisión de información desde un punto en el espacio y el tiempo hasta otro. La cantidad de información sobre un acontecimiento está estrechamente relacionado con su probabilidad de ocurrencia. Los mensajes que contienen símbolos de gran probabilidad de ocurrencia o indican muy poca incertidumbre en el resultado, conducen relativamente poca información. Por el contrario, los que contiene símbolos con baja probabilidad de ocurrencia conducen relativamente grandes cantidades de información. [1]

El objetivo de esta práctica es aplicar la teoría de la información y la medición del contenido de información con la ayuda del software Visual Basic Applications de Excel, en el cual se va a realizar los cálculos estadísticos de una cadena de caracteres ingresada por el usuario, y conocer tanto su probabilidad de ocurrencia, cantidad de información y su Entropía.

Objetivos

* Determinar la entropía de una secuencia de caracteres
* Obtener la cantidad de información de una cadena de caracteres.
* Encontrar las probabilidades de los caracteres.

Marco teórico

La teoría de la información, conocida también como teoría matemática de la comunicación o teoría matemática de la información, es una propuesta teórica presentada por Claude E. Shannon y Warren Weaver a finales de la década de los años 1940. Esta teoría está relacionada con las leyes matemáticas que rigen la transmisión y el procesamiento de la información y se ocupa de la medición de la información y de la representación de la misma, así como también de la capacidad de los sistemas de comunicación para transmitir y procesar información. La teoría de la información es una rama de la teoría de la probabilidad y de las ciencias de la computación que estudia la información y todo lo relacionado con ella: canales, compresión de datos y criptografía, entre otros.

Una medida clave de la información en la teoría es conocida como entropía, la que usualmente se expresa como el número promedio de bits necesarios para almacenamiento o comunicación. La entropía cuantifica la incertidumbre involucrada al encontrar una variable al azar.

Aplicaciones de tópicos fundamentales de la teoría de la información incluyen compresión sin pérdida de datos (ej. Archivos ZIP), compresión de datos con pérdida (ej. MP3s), y codificación de canal (ej. Para líneas DSL). El campo está en la intersección de las matemáticas, estadística, ciencias de la computación, física, neurobiología e ingeniería eléctrica.

La cantidad de información para un símbolo según Shanon puede ser hallada como:

Donde pi es la probabilidad para el símbolo sí. La información aportada por un símbolo que es la concatenación de otros dos es la suma de las informaciones de ambos símbolos.

Así mismo la entropía de un sistema de información esta expresado como:

Donde H es la entropía, las Pk son las probabilidades de que aparezcan los diferentes códigos y m el número total de códigos.

Procedimiento

Se inicia creando la interfaz gráfica en la cual se crea una ventana con 3 botones, 2 cajas de texto y diferentes etiquetas; en la caja de texto 1 se ingresa el mensaje de mínimo 40 caracteres; la ventana muestra la cantidad de símbolos actuales que tiene el mensaje.

El botón de calcular realiza el conteo, las probabilidades, y la entropía de los símbolos; el botón de limpiar borra los datos que se tiene del mensaje actual, el botón de salir cierra la ventana del cálculo de entropía.

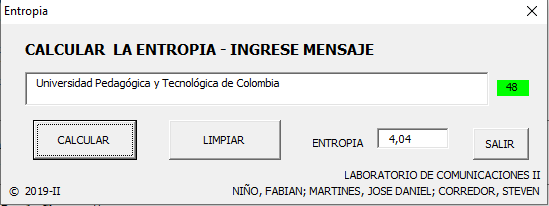


Figura . Ventana de visualización calculo entropía de un mensaje

Para escribir en las celdas de la hoja actual de Excel se usa la función Range como se puede observar en la figura 2, aquí se tienen los valores que van a tener los títulos de las columnas de cada uno de los valores que se van a hallar.

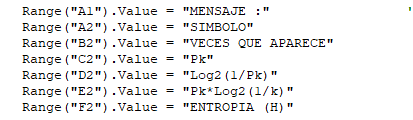


Figura . Comando de asignación de títulos a celdas

Para contar los símbolos de la cadena se utilizó un condicional por símbolo, y la función Mid(cadena, i, 1) y se va recorriendo el vector en el que se guardó el mensaje con un for, a su vez que sus resultados se guardan en un vector.

En la figura 3 se puede observar la clasificación de cada símbolo en la cual se va recorriendo el vector cadena el cual lleva el mensaje ingresado, tomando un dato a la vez; antes de clasificarlos se convierte el carácter a minúscula, si es posible, y si no cae en ninguna clasificación se ubica en un símbolo NAN.

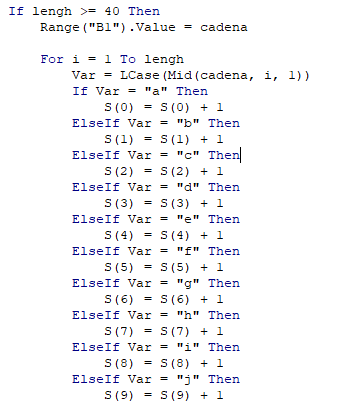


Figura . Asignación de cada símbolo del mensaje a una variable

Al clasificarlos y guardar la cantidad de símbolos se puede obtener la probabilidad de cada símbolo por la cantidad de símbolos que tiene el mensaje. Teniendo su probabilidad se procede a calcular la cantidad de información con la Ec 1, se calcula la entropía parcial de cada símbolo, guardando estos datos en vectores, la entropía del mensaje se halla usando la Ec. 3. El código implementado para esto se puede observar en la figura 4.

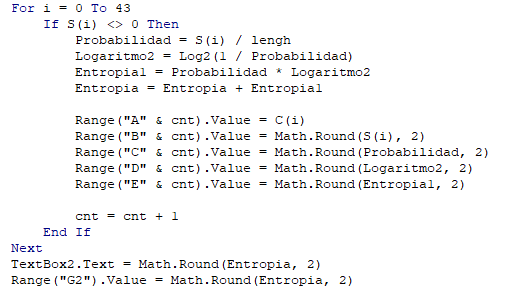


Figura . Creación de funciones de probabilidad y entropía y su asignación a las celdas

El cálculo del logaritmo en base 2 se usa una función a la cual se va a llamar para cada símbolo; como VBA no posee logaritmo en base 2, se utiliza la función



Figura . Función logarítmica en base 2

En la figura se muestra el código implementado para contar y mostrar la cantidad de símbolos que se ingresan.

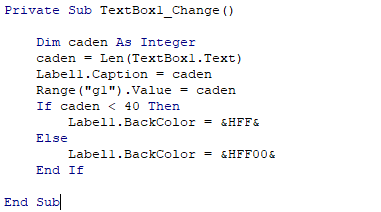


Figura . Comando de conteo de símbolos que se visualiza en etiquetas

La tabla en Excel después de calcular un mensaje ingresado en la interfaz aparecerá en la hoja de Excel como lo muestra la figura 7, solo los símbolos que aparecen en el mensaje se van a mostrar en la tabla, ya que para los demás su probabilidad es cero.

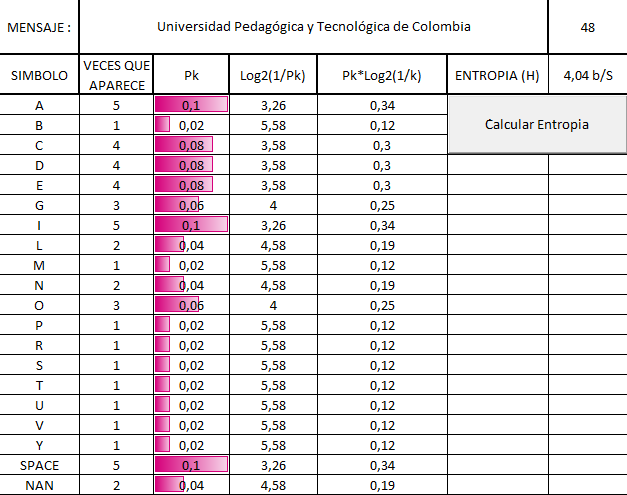


Figura . Tabla de cálculo de la entropía, símbolos y calidad de la información del mensaje

El botón de limpiar tiene como objetivo reiniciar las variables y limpiar las celdas de Excel, el código implementado para esta función se muestra en la figura 8.

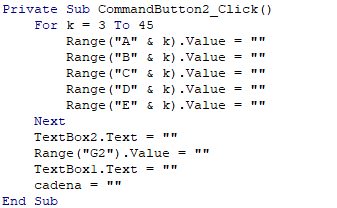


Figura . Comando de acción botón limpiar

Conclusiones

La cantidad de información recibida del conocimiento de ocurrencia de un evento es inversamente proporcional a la probabilidad de su ocurrencia.

La teoría de la información permite entender, qué es la información, cómo se mide, cuánto ocupa y cuál es la capacidad mínima de un canal de transmisión para poderla transmitir.

Referencias

1. Stremler, Ferrel G. Introduccion to comunication systems. Addison-Wesley Publishing Company, 2ª Edición. México, 1982.
2. Haykin, Simon. Sistemas de comunicación. Limusa Wiley, 1ª Edición. México, 2002.
3. Abramson N. (1963), Teoría de la Codificación. McGraw Hill.