

# TEORIA DE LA INFORMACION Y LA CODIFICACION

**Resumen:** Esta práctica tiene como objetivo obtener la entropía de un texto, la probabilidad de los caracteres (incluyendo puntos, comas etc..).

## I. INTRODUCCION

El objetivo de esta práctica es entender la teoría de la información, el uso de Visual Basic en Excel para la lectura de una base de datos, en este caso compuesta por una cadena de caracteres, y aprender el lenguaje de Visual para aplicaciones VBA.

## II. 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.

## III. MATERIALES

- Visual Basic

## IV. MARCO TEORICO

Teoría de la información Fue desarrollada por Claude E. Shannon para encontrar los límites fundamentales en la compresión y almacenamiento confiable y comunicación de datos. Se ha ampliado para encontrar aplicaciones en muchas otras áreas, incluyendo inferencia estadística, procesamiento del lenguaje natural, criptografía, otras redes diferentes a las redes de comunicación como en

neurobiología, la evolución y función de códigos moleculares, selección de modelos en ecología, física térmica, computación cuántica, detección de plagiarismo y otras formas de análisis de datos.

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 Shannon puede ser hallada como:

$$I(S_i) = \log_2 \left( \frac{1}{P_i} \right)$$

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

$$I(S_i S_j) = I(S_i) + I(S_j)$$

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

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

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

## V. PROCEDIMIENTO

### Parte 1

Realizar una interfaz gráfica en visual Basic en Excel, por medio del uso de un botón que de acceso a la misma y que permita ingresar una cadena de caracteres “mensaje” de mínimo 40 caracteres, el cálculo de la probabilidad de los mismos, la cantidad de información “caracteres” y la entropía del mensaje, se debe visualizar en la hoja de Excel en una columna cada carácter y al frente la probabilidad del mismo. Realizar La toma de datos de dicho Excel en Matlab.

### BIBLIOGRAFICA

[1] Abramson N. (1963), Teoría de la Codificación. McGraw Hill.

[2] Jair N. Bernal. TEORIA DE LA INFORMACION Y CODIFICACION

[3] Sistemas de Comunicación Digitales y Analógicos - Leon W. Couch - 7ed