

# Proyecto 2 - Codificación para fuentes con memoria

Sistemas de Comunicaciones Digitales I - 2024

## Objetivos

1. Investigar en la bibliografía lo siguiente
  1. Proceso de Markov o cadena de Markov y sus propiedades
  2. Fuente de Markov
  3. Entropía Condicional
  4. El algoritmo LZ77 de codificación para fuentes con memoria
2. Realizar las siguientes actividades, programando en lenguaje Python
  1. Desarrollar una rutina de análisis que, tomando como entrada un archivo de texto utf-8, genere una tabla de frecuencias de palabras condicionales a las últimas  $N$  palabras encontradas, donde  $N$  es un parámetro.
  2. Realizar un modelo de fuente discreta de Markov que, a partir de una tabla de frecuencias condicionales, genere texto aleatorio.
  3. Realizar una implementación del algoritmo LZ77, tanto codificación (compresión) como decodificación (descompresión), que tome como entrada una cadena de caracteres y produzca una salida binaria representada por una cadena de unos y ceros. Al emitir literales utilizar el código utf-8 (la salida sigue siendo una representación binaria con caracteres unos y ceros, pero según dicho código).
  4. Evaluar el índice de compresión de la implementación de LZ77 contrastando la cantidad de *bits* (bytes multiplicado por ocho) ocupados por la cadena original codificada en utf-8 con la cantidad de *bits* (unos y ceros) utilizados por el algoritmo de LZ77 para representarla. Para ello utilizar la fuente de Markov desarrollada, con una tabla de frecuencias obtenida del análisis de un texto descargado de [Proyecto Gutenberg](#)

## Entregables

La entrega se realizará mediante un repositorio git público alojado en GitHub y creado como Fork de este repositorio ([https://github.com/fmirandabonomi/scd\\_2024-proyecto\\_2](https://github.com/fmirandabonomi/scd_2024-proyecto_2)). Debe incluir el código desarrollado y un informe escrito.

El informe escrito debe contener las siguientes partes

- *Título*
- *Autor*

- *Resumen.* Luego de leer el resumen y sin necesidad de ver el resto del trabajo, el lector debe ser capaz de comprender en forma general *sobre que tema trata el trabajo, que se hizo, como se hizo y los principales resultados obtenidos.*
- *Introducción.* Luego de leer la introducción el lector debe comprender en detalle sobre que tema trata el trabajo y cual es el alcance de la investigación y el desarrollo. *Deben utilizarse citas bibliográficas en formato APA para indicar las fuentes consultadas.*
- *Metodología.* Aquí se expondrán los resultados de la investigación bibliográfica del primer objetivo. Pueden haber varias secciones. Pueden utilizarse tablas, figuras, listados, ecuaciones, etc. Cada elemento debe tener su correspondiente referencia y título o epígrafe. Deben citarse fuentes de las ideas expuestas (salvo ideas fundamentales que son resultados ampliamente conocidos, desarrollos realizados por el autor del informe en forma independiente y resultados originales). Si se incluyen figuras u otro inserto tomado de alguna fuente bibliográfica debe atribuirse correctamente. Toda figura, tabla, listado o ecuación debe ser mencionada en el texto utilizando la correspondiente referencia.
- *Resultados y discusión.* Aquí se explican los desarrollos realizados en el objetivo 2. Se da suficiente detalle para que el lector comprenda *que se hizo y como funciona* pero **no se duplica el código fuente**, solo las *ideas fundamentales* y el *diseño de alto nivel* del mismo. Se discute que significan los resultados obtenidos. Iniciar la sección con un pequeño resumen de los resultados obtenidos.
- *Conclusiones.* En esta sección se explica lo aprendido durante la investigación. Todo lo que se diga debe ser *basado en los resultados obtenidos* y debe ser *objetivo* (no una opinión). Un lector que leyó el resto del trabajo debe ser capaz de llegar a las condiciones obtenidas (o contradecirlas) usando solo lo expuesto y un proceso de razonamiento.
- *Referencias.* En esta sección se listan, en formato APA, las referencias bibliográficas consultadas. Todas las fuentes listadas deben haber sido utilizadas en la elaboración del informe (deben haber sido citadas en el texto).

## Evaluación

El proyecto será evaluado en base al código desarrollado (50% de la calificación) y el informe presentado (50% de la calificación).