

Ejercicio AvCont-4(b)

- Enunciado – parte b:

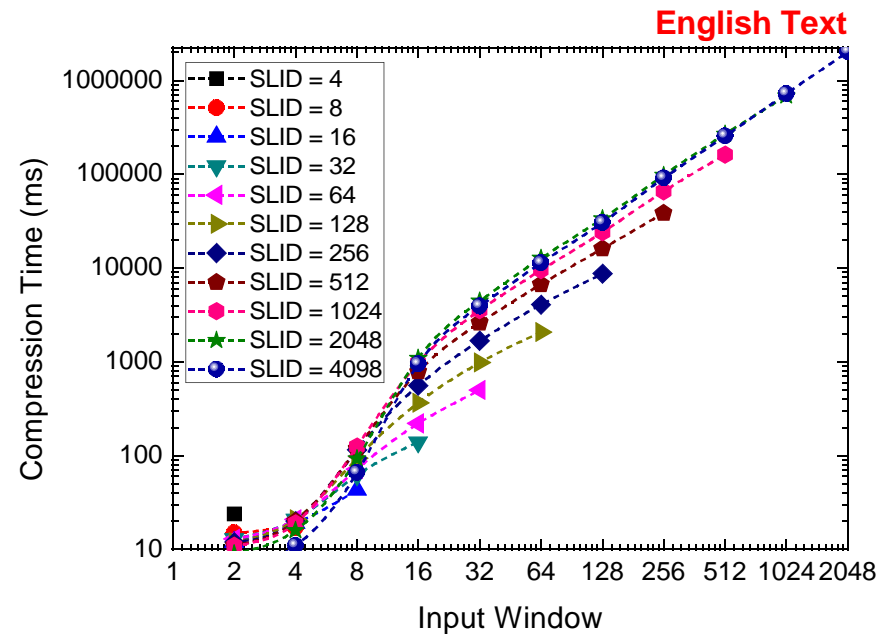
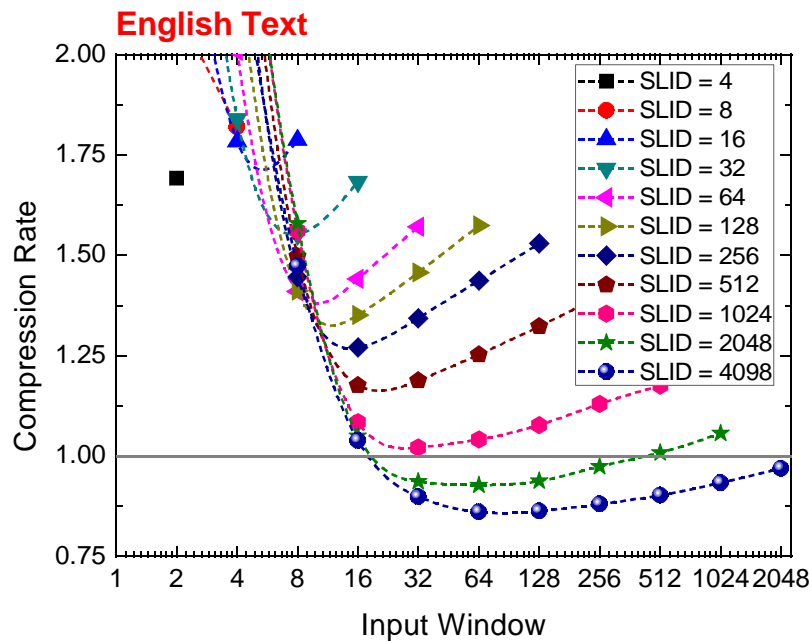
Después de haber implementado en Java un compresor/descompresor de datos binarios mediante el algoritmo LZ-77 que cumpla las especificaciones indicadas en la parte (a) del ejercicio, analizaremos ahora la **capacidad de LZ-77 para comprimir texto**.

- c) Haced las modificaciones necesarias en el programa anterior para que sea capaz de **leer ficheros de texto, generar el código ASCII (de 8 bits) de cada carácter del texto y devolverlo como una *String binaria* procesable por vuestros métodos de compresión /descompresión en LZ-77**. Conseguir también que el programa **calcule el tiempo invertido en comprimir** los datos. Comprobad el correcto funcionamiento del programa; es decir, que es capaz de comprimir texto en LZ-77 y de recuperarlo tras la descompresión. Ayuda: Utilizad los métodos de la clase *txtReader* que se os proporciona Ayuda: Medid el tiempo mediante *System.nanoTime()*. Ayuda: Se os proporcionan los archivos “hamlet_short.txt” y “quijote_short.txt”, que contienen dos fragmentos de texto en inglés y en español, respectivamente.
- d) Comprimid el archivo “hamlet_short.txt” con distintos valores de M_{des} y M_{ent} entre 4 y 4096 y analizad el factor de compresión y el tiempo invertido para conseguirlo. ¿Cuál es el mejor factor de compresión que obtenéis y con qué valores de M_{des} y M_{ent} ? ¿Cómo varia (cómo escala) el tiempo de cálculo necesario al aumentar M_{des} y M_{ent} ? ¿Qué combinación de M_{des} y M_{ent} elegiríais?
- e) Comprimid ahora el archivo “quijote_short.txt” y analizad para qué combinación de M_{des} y M_{ent} se obtiene el mejor factor de compresión. ¿Es el mismo que en el caso anterior? Proponed varias razones que expliquen esta diferencia.

(cada apartado correcto cuenta como uno de los ejercicios hechos anteriormente)
(podéis trabajar por grupos)

Ejercicio AvCont-4(b)

- Solución – parte b:



Las repeticiones propias del texto facilitan la compresión

Requiere:

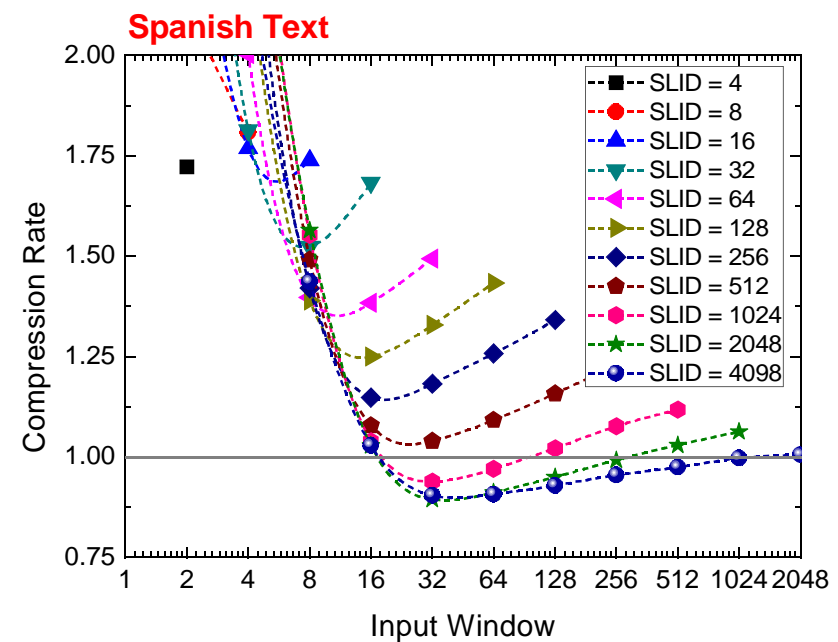
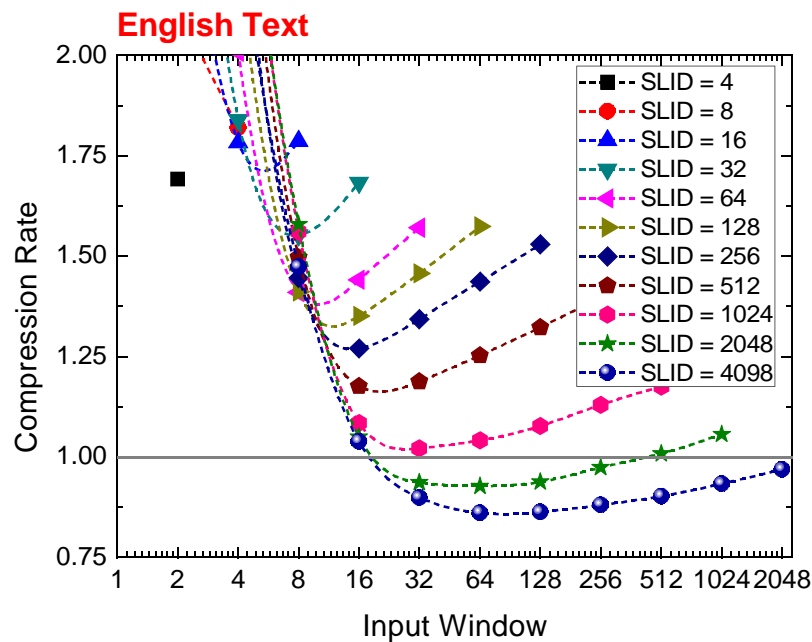
M_{ent} que incluya varios caracteres

M_{des} larga que aumente la probabilidad de coincidencia

Algoritmo de compresión de escalado lineal $O(N)$

Ejercicio AvCont-4(b)

- Solución – parte b:



La eficiencia de la compresión depende de las características del texto