

Práctica 2

La transformada de Burrows-Wheeler (BWT) convierte una cadena cualquiera en otra que tiene mucha mayor frecuencia de símbolos iguales contiguos, y además este proceso es reversible. Una de sus utilidades más comunes es en la compresión de texto, ya que se utiliza junto con Move-To-Front y Huffman para obtener el popular compresor `bzip2`.

La forma más común de calcular la BWT es ordenando adecuadamente las rotaciones de la cadena de entrada. Esto se puede hacer directamente con un coste de $O(n^2 \log n)$, donde n es la longitud de la cadena, o mucho más eficientemente con coste en tiempo cuasilineal utilizando vectores de sufijos.

Hay que tener en cuenta lo siguiente

- Cualquier fragmento de código o idea para el que se utilicen fuentes externas debe ser identificado y dichas fuentes citadas.
- En la práctica hay que hacer como mínimo lo que se especifica a continuación, pero la tarea completa no está limitada.
- En la evaluación de esta práctica se tendrán especialmente en cuenta dos aspectos: los detalles de implementación de los vectores de sufijos y la eficiencia de la implementación realizada.

1.1. ¿Qué hay que hacer?

1. Implementar de la forma más eficiente posible la estructura de datos de vectores de sufijos. Deben explicarse cada uno de los detalles de la implementación que hayan permitido alcanzar dicha eficiencia.
2. Implementar la transformada de Burrows-Wheeler utilizando vectores de sufijos, intentando llegar a coste cuasilineal en tiempo (Nota: La eficiencia de esta implementación es apreciable para cadenas muy largas. No es necesario, aunque está permitido, investigar otras opciones distintas a los vectores de sufijos).
3. Implementar el algoritmo de compresión Move-to-Front.
4. Probar todos los algoritmos implementados con un número suficiente de datos interesantes y de suficiente tamaño.

5. Utilizando alguna implementación pública del código de Huffman (o una propia), combinar BWT, Move-To-Front y Huffman para obtener un compresor similar a `bzip2`.

1.2. Entrega

Deberá entregarse **hasta el 21 de enero de 2020**.