**Me falta hacer la portada.**

**Me disculpo por el retraso, se me olvidó que tenía que trabajar el jueves y tuve unos cuantos problemas el fin de semana.**

**Me falta hacer revisión de ortografía.**

**Las líneas en rojo son cosas que no van en el informe final, son comentarios y/o observaciones mías sobre lo que escribí.**

**No sé si existe algún estándar de formato, no sé si tiene que ser doble espaciado o si tengo que usar algún tamaño de letra específico.**

**Por el apuro lo hice en Word. Tengo intenciones de hacer el informe final del ramo en LaTeX.**

## Motivación

El problema de los “optimal prefix free codes” (en adelante OPFC) consiste en encontrar para cada elemento perteneciente a un set símbolos una codificación optima de tal forma que ningún código sea prefijo de otro. Esta codificación es óptima en el sentido de que el tamaño del código de cada elemento depende de la frecuencia de los símbolos.

Este problema ha sido ampliamente estudiado a lo largo de la historia y hay varios resultados interesantes, siendo su implementación más conocida (y enseñada) la codificación de Huffman.

Está demostrado que el tiempo óptimo para determinar los OPFC de un set de símbolos, dada una lista de frecuencias y usando una cola de prioridad es $\theta(nlogn)$ en el modelo de comparaciones. Es decir no existe una solución que de mejores resultados en tiempo. Sin embargo Van Leuween demostró que el cálculo de la codificación de Huffman podía llevarse a cabo en tiempo O(n) si la lista de frecuencias era entregada inicialmente ordenada.

Esto nos da como intuición que la cota $\theta(nlogn)$ es más que nada porque calcular los OPFC implica (ya sea directa o indirectamente) ordenar la lista de frecuencias. A partir de esto surgen dos interrogantes, que son las que se pretenden contestar durante la realización del trabajo:

1. ¿Es posible computar los OPFC en tiempo mejor que $\theta(nlogn)$ al sacar ventaja de ciertos “casos fáciles” específicos?
2. ¿Es posible calcular los OPFC sin errores haciendo solamente un ordenamiento parcial de la lista de frecuencias?

## Discusión

El problema de los OPFC se puede enunciar de una manera un poco más formal como:

"Sea T un texto formado por elementos pertenecientes a un lenguaje L. Se sabe que cada palabra p perteneciente a L ocupa una cantidad k de bits. Dada una lista de frecuencias de aparición de cada palabra de L en T, se requiere encontrar una nueva codificación para cada palabra de este lenguaje de forma que se genere un nuevo texto T' a partir de T que ocupe menos espacio." (Sé que puedo ser aún más formal al especificar mejor el lenguaje o decir cuánto es k (dado el tamaño del lenguaje).)

La codificación de Huffman resuelve exactamente este problema en, como ya se dijo, en $\theta(nlogn)$. Pero las preguntas propuestas apuntan a que se puede lograr algo mejor, si consideramos que tenemos acceso a más información.

Al recibir la lista de pesos, se puede considerar el reconocer casos que sean fáciles de resolver. Por ejemplo:

1. Si la lista de frecuencias viene ordenada: revisar que una lista viene ordenada toma tiempo lineal, y luego solo hay que aplicar Van Leeuwen en vez del algoritmo clásico.
2. Si la frecuencia de todas las palabras es la misma: revisar que todos los elementos de una lista son los mismos toma tiempo lineal, y construir los OPFC dado eso también.
   (Me habría gustado poner más, pero no tuve mucho tiempo para pensar.)

También algo a lo que se le puede sacar ventaja es saber que existen algoritmos adaptivos de ordenamiento. Estos algoritmos pueden ordenar en tiempo menor a la cota $\theta(nlogn)$ aprovechándose del hecho que fragmentos de un arreglo pueden venir parcialmente ordenados. Entonces ordenando de forma adaptiva y después haciendo el cálculo de los OPFC se puede ver intuitivamente que se puede hacer algo mejor. (Quizá debería ahondar un poco más en este tema)

Hasta el momento se ha asumido que es necesario ordenar completamente la lista de frecuencias para poder hacer el cálculo de los OPFC. Sin embargo existe la intuición de que esto puede ser hecho directamente sobre un arreglo no desordenado, si no que parcialmente ordenado.

La intuición de por sí es correcta, por ejemplo si se tiene la siguiente lista de frecuencias:

| 2 | 1 | 4 | 5 | 8 | 7 | 11 | 10 |
|---|---|---|---|---|---|---|---|

Sobre esta lista da lo mismo si se ordena o no, ya que Van Leeuwen extrae inicialmente los dos elementos iniciales de la lista (asumiendo que son los mínimos), cosa que en este caso se cumple a pesar de que no están ordenados. (No sé si será necesario que aquí explique Van Leeuwen con más detalle).

Este caso de por sí es muy simple (y muy conveniente), pero generalizar esto no es trivial, y más aún, no se está mostrando que se puede calcular los OPFC sobre un arreglo no desordenado, solo se está mostrando que aplicar el algoritmo de Van Leeuwen sobre esta entrada específica no tiene efecto en el resultado final. Pero en realidad en ningún caso se debería aplicar Van Leeuwen sobre un arreglo desordenado.

Esto de hecho, refina la pregunta planteada anteriormente y podemos reformularla como:

Sabiendo que es posible calcular los OPFC sobre arreglos parcialmente ordenados. ¿Es posible modificar el algoritmo de Van Leeuwen para que funcione sobre estos arreglos? ¿Al recibir un arreglo desordenado, es posible calcular los OPFC de forma que al terminar de calcularlos el arreglo no quede ordenado?

## Objetivos

**General:** Encontrar una respuesta formal a las preguntas planteadas.

Se nos indicó que el objetivo general debería ser solo una frase, pero siento que lo que yo escribí es demasiado genérico.

**Específicos:**

1. Hacer un estudio teórico sobre las preguntas planteadas para lograr un mejor entendimiento del tema y proponer una mejor solución.
2. Definir los algoritmos necesarios para poder responder las preguntas planteadas
3. Implementar los algoritmos definidos y testearlos de forma de obtener datos empíricos y hacer que el experimento sea repetible.

(Debería poner referencias acá, pero lo hice casi todo de memoria. De Van Leeuwen me acorde mirando la presentación que hice el año 2013)