

Tarea 1: Esquemas de Decodificación Instantánea - INFO145

Diseño y Análisis de Algoritmos.

Académico: Héctor Ferrada.
Instituto de Informática, Universidad Austral de Chile.

Abril 27, 2022

Entrega. Debe subir todo a siveduc en un archivo *t1Apellidos.zip* (3 a 4 integrantes por grupo), con su implementación y un informe claro y conciso. Fecha de entrega: **Jueves 19 de Mayo.**

Informe. El informe debeser claro, objetivo y detallado, este debe contemplar: Resumen, Introducción, metodología, experimentación y conclusiones. Se debe introducir adecuadamente el contexto del problema, explicar la lógica que usó en sus implementaciones y la justificación de ello (p. ej. explicar por qué no uso otra alternativa más eficiente si la hubiese), detallar la experimentación realizada, los gráficos de sus resultados y las conclusiones de su trabajo; con todo esto debe validar o rechazar sus hipótesis iniciales justificando detalla y objetivamente.

Rendimiento en Algoritmos de Codificación/Decodificación

El objetivo de este trabajo es adquirir un mayor dominio en el análisis de algoritmos y de estructuras de datos para resolver un mismo problema. Se evalúa que usted pueda llevar a la práctica la teoría vista en clases, implementando los algoritmos de **Shannon Fano** y de **Huffman** para codificar símbolos, chequeando el performance empírico de sus soluciones implementadas.

Codificación de secuencias de símbolos. El problema a resolver es el siguiente. Dado un texto $T[1..u]$, construido sobre un alfabeto $\Sigma[1..\sigma]$, se desea construir un bitstream (secuencia de bits) $B[1..n]$, por medio de los esquemas de decodificación instantánea ya introducidos, que represente unívocamente a T . Deberá estimar el **espacio exacto en RAM** que requieren sus estructuras construidas y ejecutar experimentos que midan la **velocidad de decodificación de sus implementaciones** a fin de incluir un buen análisis.

Algoritmo de Shannon Fano. Corresponde al mismo algoritmo estudiado en clases, el cual deberá implementar para construir su representación $F[1..n]$. Debe además **proponer e implementar** un algoritmo de decodificación, explique su nuevo algoritmo y entregue el análisis del tiempo de ejecución en función del tamaño de la entrada n y de σ si es necesario.

Algoritmo de Huffman Canónico. Corresponde al mismo algoritmo estudiado en clases, el cual deberá implementar para construir su representación $H[1..m]$; además del algoritmo de decodificación que también fue entregado en clases.

Al construir las estructuras debe considerar que el espacio total de cada método debe incluir al bistream junto a todas las estructuras de datos necesarias para ejecutar la decodificación. Para el caso de cada bistream $F[1..n]$ y $H[1..m]$, puede utilizar arreglos booleanos para representar los bits. Considere para sus análisis que, en el modelo RAM, se considera tiempo constante la lectura/escritura de $O(\log n)$ bits; por ejemplo, esto se ve en el algoritmo de decodificación de Huffman canónico al formar el entero N en un solo paso; a pesar de que, quizá, en su implementación lo haga bit a bit y no los $h = O(\log n)$ bits de un solo paso (esto si no dispone de alguna librería adicional que le permita leer/escribir h bits desde/en un entero).

Experimentación. Descargue los siguientes archivos para ejecutar sus experimentos:

- english.100MB.
- dna.100MB.

Utilice estos archivos para ejecutar experimentos con texto en inglés y con secuencias de ADN. Cada uno de estos archivos hace el papel del $T[1..u]$ de entrada. Para cada T , construya sus estructuras codificadas F y H , luego realice al menos los siguientes experimentos —desde luego, puede agregar más experimentos si estima que es necesario para su análisis y presentación clara de los resultados.

1. **Tiempo de extracción de prefijos del texto.** Para diferentes posiciones aleatorias del texto $k \leq u$, extraiga $T[1..k]$ desde sus representaciones codificadas.
2. **Tiempo de extracción de segmentos aleatorios del texto.** Previo a la ejecución de experimento, **deberá proponer nuevos algoritmos** que le permitan recuperar **eficientemente** cualquier segmento $T[i..j]$, con $1 < i \leq j \leq u$, sin la necesidad de comenzar la decodificación desde el inicio de su representación codificada (evitando hacer lo que se hizo en el experimento anterior). Explique sus nuevos algoritmos así como las estructuras de datos adicionales incluidas para poder llevar a cabo su nueva metodología.
3. **Tasa de compresión y espacio utilizado.** Determine el espacio utilizado por sus implementaciones, considerando la cantidad de memoria que requieran las estructuras auxiliares de las que dependan sus algoritmos. Estime y compare el tamaño de los archivos comprimidos que, hipotéticamente, generarían sus implementaciones.

Grafique los resultados de sus experimentos para luego **analizar y concluir sobre los eventos**. Comente también sobre la forma y las condiciones en las cuales llevó a cabo la experimentación, incluyendo detalles que permitan reproducir sus resultados.

Evaluación Su trabajo será evaluado con dos notas:

- **Nota de informe.** Una nota por el informe que pondera un 60 % del trabajo. En esta se evaluará el trabajo teórico y el análisis que ha realizado. Además de la estructura, presentación y calidad de su escrito; la cual también incluye la presentación de los resultados obtenidos y sus conclusiones.
- **Nota de implementación.** Una nota por las implementaciones¹ pedidas que pondera un 40 % del trabajo. No solo se espera que entregue códigos correctos, sino que además estén ordenados, sean modulares y que estén debidamente documentados. Por favor incluya un archivo README.txt con las instrucciones de ejecución y dependencias si las hubiese.

¹Puede utilizar C++, Python o Java. Aunque siempre es recomendable que en este tipo de investigaciones se emplee C++, su nota no se verá afectada por la elección del lenguaje de programación

Se adjunta un archivo zip con algunas funciones e instrucciones adicionales que pudiesen ser de utilidad. Cualquier consulta puede realizarla al auxiliar de la asignatura a través de su correo institucional (jose.vasquez01@alumnos.uach.cl) o por medio de Discord.