



Estructuras de Datos y Algoritmos Avanzados (2025-2)

Laboratorio 3: String matching

Profesor: José Fuentes Sepúlveda

Ayudante: Luciano Hernán Argomedo Solís

Objetivos

Los objetivos del laboratorio son:

- Entender el código de algoritmos y estructuras de datos de complejidad media.
- Estudiar la documentación de la biblioteca estándar de C++ y utilizar algoritmos ya implementados.
- Evaluar experimentalmente implementaciones de algoritmos.

Ejercicios

1. Implementar u obtener la implementación de un *FM-index*, el cual debe implementar una operación de búsqueda que retorne la cantidad de veces que se repite un patrón en el texto sobre el que se construye la estructura. En caso de utilizar código que no sea de autoría propia, citar la fuente y comentar brevemente el código, explicando las componentes más importantes del mismo.
2. Implementar u obtener la implementación de un algoritmo de búsqueda de patrones en cadenas de texto, el cual debe retornar la cantidad de repeticiones de un patrón en un texto. El algoritmo en cuestión puede ser *Knuth-Morris-Pratt*, *Boyer-Moore* o *Rabin-Karp*. En caso de utilizar código que no sea de autoría propia, citar la fuente y comentar brevemente el código, explicando las componentes más importantes del mismo.
3. Evaluar experimentalmente el desempeño de ambas alternativas. Para esto, realizar un experimento donde se mida el efecto del largo del texto en los tiempos de ejecución, y otro experimento para medir el efecto del

largo del patrón. Reportar los resultados a través de gráficos y tablas. Se deben utilizar textos del repositorio *Pizza&Chili*, cuya dirección es <http://pizzachili.dcc.uchile.cl/texts.html>¹. Se deben hacer experimentos sobre mínimo dos textos y estos deben tener un peso mínimo de 200 MB y como mínimo uno de los dos debe pesar más de 1 GB. Adicionalmente, reportar el consumo de memoria de cada una de estas alternativas.

4. En base a los resultados obtenidos en el ejercicio anterior, además de las complejidades obtenidas de cada alternativa a través del análisis asintótico, argumentar brevemente en qué escenarios conviene utilizar una alternativa o la otra. Tenga en cuenta no solamente el costo de la búsqueda, sino que además de la construcción, en los casos pertinentes.

Observación

Normas de entrega

Antes del lunes 10 de noviembre a las 23:59, se deben enviar todos los ejercicios resueltos mediante la plataforma CANVAS.

Se debe entregar:

- Archivo PDF correspondiente a un breve informe acerca del boletín. La portada debe indicar el nombre completo y número de matrícula.
- Todos los ficheros del código fuente dentro de un fichero comprimido.
- **IMPORTANTE:** Los archivos debe llamarse `apellido_nombre_03.formato`
- *Por favor, subir en Canvas el archivo comprimido y el informe separadamente.*

¹ Algunos de los ficheros que se encuentran en esta web no están codificados en ASCII, por lo que puede ser necesario realizar un preprocessado para ajustar la entrada