



Estructuras de Datos y Algoritmos Avanzados (2023-2) **Laboratorio 3: String matching**

Profesor: José Fuentes Sepúlveda

Ayudante: Oliver Brito Alarcón

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 *Suffix Array*, 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. Se sugiere utilizar el código dispuesto por el ayudante en esta tarea. 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 *Boyer-Moore-Horspool*. 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 puede utilizar texto de origen natural o sintético. En el repositorio *Pizza&Chili*, cuya dirección es <http://pizzachili.dcc.uchile.cl/texts.html>¹, se encuentran disponibles diversos ejemplos y generadores que se pueden utilizar en esta experimentación. Adicionalmente, reportar el consumo de memoria de cada una de estas alternativas. Se recomienda hacer uso de las cuentas en el servidor chome, ya que en la carpeta `home/datasets/` se encuentran los textos de *Pizza&Chili* y algunos otros.

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 viernes 6 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 y . La portada debe indicar el nombre completo y número de matrícula.
- Todos los ficheros del código fuente (solo los archivos `.h` y `.cpp`) junto a un archivo `README.txt` con sus instrucciones de compilación correspondiente. Estos archivos deben subirse al servidor asignado para el curso, dentro de una carpeta de nombre *Boletin_03*.
- **IMPORTANTE:** Los archivos debe llamarse *apellido1_nombre_03.formato*
- *Por favor, subir en Canvas el archivo comprimido y el informe separadamente.*

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