

**Tarea 1: Experimentaciones sobre Búsqueda de Patrones en Textos**

(Fecha de entrega: [2021-04-01 Thu])

El propósito principal de esta tarea es programar varios algoritmos de búsqueda de patrones en texto, y comparar sus rendimientos con instancias realistas. Se pone énfasis en que se pide que se propongan hipótesis, diseñen experimentos (simples) y prueben o rechacen esas hipótesis. Pueden hacer esta tarea en grupos de uno a cuatro personas, pero se recomienda hacerla en grupos de dos.

a. Programación de algoritmos de búsqueda de patrones:

- a) Encontrar una implementación del (o programar el) algoritmo de búsqueda de patrones Knuth, Morris and Pratt, **KMP**.
- b) Encontrar una implementación del (o programar el) algoritmo de búsqueda de patrones Boyer and Moore, **BM**.
- c) (Para los grupos de 3 o 4 personas) Encontrar una implementación del (o programar el) algoritmo de búsqueda de patrones Boyer, Moore and Sunday, **BMS**.
- d) (Para los grupos de 4 personas) Encontrar una implementación del (o programar el) algoritmo de búsqueda de patrones Boyer, Moore and Horspool, **BMH**.

b. Encontrar ejemplos reales de textos  $T$  y patrones  $P$  tales que crean que

- a) **KMP** será más rápido que los otros;
- b) **BM** será más rápido que los otros;
- c) (Para los grupos de 3 o 4 personas) **BMS** será más rápido que los otros;
- d) (Para los grupos de 4 personas) **BMH** será más rápido que los otros;

c. Experimentación:

Ejecutar todos los algoritmos sobre todos los textos. Es esperable y completamente aceptable que algunas de sus hipótesis no sean verificadas.

d. Redactar un reporte de 2 a 4 páginas (una hoja contiene dos páginas) con

- a) una lista enumerada de los nombres completos de los miembros del grupo, uno por línea;
- b) Descripción corta, código (y referencias exactas si creado por otros) para cada implementación;
- c) Referencias y descripciones cortas de los textos y patrones usados;
- d) una matriz de 2x2, 3x3 o 4x4 con los resultados experimentales, en cantidad de comparaciones de caracteres o en tiempos de ejecución. (Las filas y columnas vendrían a ser los ejemplos y los algoritmos);

- e) Un párrafo (potencialmente corto) para cada experimento (p.ej. 4 si en un grupo de 2, 9 si en un grupo de 3, 16 si en un grupo de 4) describiendo el resultado del experimento, y en particular indicando claramente si la hipótesis (algoritmo  $A$  mejor que algoritmo  $B$ ) fue confirmada por el ejemplo o no;
- f) Menciones de todas las fuentes de información recibida fuera de los co-autores de la tarea. Ojo que si esta tarea es asignada a un alumno para un examen oral, dicho alumno deberá explicar la tarea completa, no solamente la parte en la que se haya enfocado.

Observaciones:

- a. La tarea se puede hacer solo (pero con todo el trabajo requerido para dos), en grupos de dos (lo recomendado) o en grupos de tres o cuatro (con una cantidad de trabajo total que aumenta cuadráticamente con el tamaño del grupo!): se recomienda **fuertemente** hacerla en grupos de dos.
- b. Pueden elegir el lenguaje de programación, pero se recomienda **Python 3**.
- c. Pueden entregar su reporte en formato **pdf** (e.g. generado via  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ ) o en formato **notebook** (formato recomendado), pero deben incluir todo el código fuente en archivos separados..
- d. Están súper bienvenidos a intercambiar fuentes de implementaciones de algoritmos, de conjuntos de textos, trucos para leer los textos en sus programas, etc. (p.ej. en el canal **tarea1** de **Discord**, para que los alumnos a quienes les toque presentar esta tarea en un examen oral puedan encontrar toda la información junta). Si usan una de tales sugerencias, agradezcan en su reporte los autores de las sugerencias, mencionando dónde se encontró la sugerencia (p.ej. “Javier Olivar en el canal Tarea1 del Discord del Curso, el [2021-03-18 Thu]” o “Francisco Brieve, comunicación personal, el [2021-03-25 Thu]”).
- e. Está sugerido **NO** mencionar afuera del grupo de trabajo (porque es parte clave de la tarea)
  - a) cuáles textos y patrones específicos se han elegido;
  - b) resultados específicos de los experimentos;
  - c) explicaciones de por qué pasa algo (en particular contrario a las expectativas/hipótesis).