

Mini-Proyecto II: Búsqueda en texto

Descripción

El problema básico de búsqueda en texto consiste en encontrar todas las ocurrencias de un patrón P de largo m en un texto T de largo n . Existen diferentes soluciones a este problema que tratan de mejorar el algoritmo de fuerza bruta (complejidad $O(nm)$) asintóticamente o en la práctica. En este proyecto se implementará como *baseline* el algoritmo de fuerza bruta. Además, se implementarán algunos de los algoritmos más conocidos para resolver el problema (Boyer-Moore, Knuth-Morris-Prat, Rabin-Karp, etc.).

Objetivos específicos

La entrega de la tarea (un informe y código fuente) debe satisfacer los siguientes objetivos específicos:

- Describir las características principales de los algoritmos seleccionados.
- Describir las decisiones de implementación más importantes.
- Evaluar experimentalmente los algoritmos seleccionados. Dado que los algoritmos son relativamente sencillos y conocidos, la carga principal del proyecto reside en esta evaluación experimental. Recomendaciones, emplear diferentes *datasets* (sintéticos vs. reales), estudiar influencia del tamaño del texto/patrón, buscar peores/mejores casos, etc.

Condiciones

- El proyecto se realizará en grupos de dos o tres estudiantes. El informe debe reflejar claramente los autores del proyecto.
- Se deben implementar y evaluar tantos algoritmos como miembros tenga el grupo, además del *baseline*. Es decir, un grupo de X estudiantes, debe implementar $X+1$ algoritmos (incluyendo entre ellos el algoritmo de fuerza bruta).
- Todas las implementaciones del grupo deben ser en el mismo lenguaje de programación y deben emplear el mismo tipo de optimizaciones.
- Los lenguajes de programación aceptados son C, C++, Java y Python.
- El *baseline* debe ser implementado entre todos los miembros del grupo. Lo mismo sucede para la evaluación experimental y para la redacción del informe.

Entrega

La entrega se realizará en la plataforma Piazza. La fecha límite para la entrega será el viernes 21 de octubre a las 23:59. Se enviará un mensaje a “Instructors” con las siguientes características:

Folder: hw2

Summary: Strings

Details: Adjuntar informe en formato pdf y código.