

Árbol de sufijos

Algoritmia para problemas difíciles

Cristian Román Morte (646564)

Aron Collados Torres (626558)

Índice

[Índice](#)

[Introducción](#)

[Ejecutables](#)

[Creación de un Árbol de Sufijos Compacto](#)

[Pruebas](#)

[Reparto de tareas](#)

[Bibliografía](#)

Introducción

Un árbol de sufijos es una estructura de datos que sirve para almacenar una cadena de caracteres con “información pre-procesada” sobre su estructura interna.

Se ha implementado un programa capaz de crear un árbol de sufijos a partir de uno o varios textos, y realizar consultas buscando en texto por un String o un patrón en particular.

Ejecutables

- ejecutar2.sh [patron] [texto]
- ejecutar2.sh [patron] [N] [t1] [t2] [tN]
- ejecutar_tests2.sh: ejecuta los tests explicados en la sección pruebas

Creación de un Árbol de Sufijos Compacto

La creación del árbol se realiza en dos etapas, en la primera etapa se recorre cada texto sobre el que se quiere hacer una consulta, se generan todos los posibles sufijos del texto, y se van añadiendo al árbol.

En la segunda etapa, una vez se tiene el árbol formado, se recorre entero, realizando una compactación de los nodos que no tengan hijos terminales.

Pruebas

Se ha implementado un generador de textos aleatorios, para realizar las pruebas. Al ejecutar la clase Test, se generan 100 textos aleatorios, de entre 150 hasta 200 caracteres, que se insertan en un árbol de sufijos. A partir de ese árbol se realiza 4 pruebas distintas:

- 100000 tests de sufijos con partes de los textos elegidas aleatoriamente
- 100000 tests de sufijos con cadenas de texto generadas aleatoriamente
- Matching de todos los textos sobre el árbol
- 100000 test de matching con cadenas de texto generadas aleatoriamente

| Prueba | Tiempo |
|-------------------------|---------|
| Creación del árbol | 1857 ms |
| Compactación del árbol | 977 ms |
| 200000 tests de sufijos | 1270 ms |

| | |
|--------------------------|-------|
| 100100 tests de matching | 96 ms |
|--------------------------|-------|

Reparto de tareas

Aron Collados: Creación y compactación del árbol

Cristian Román: Substring, matching y tests

Bibliografía

- <http://webdiis.unizar.es/asignaturas/TAP/material/4.2.sufijos.pdf>