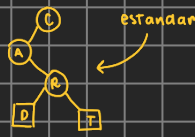


- Estructura de datos basada en árboles
- Se enfoca en buscar patrones en una colección de cadenas
- Encuentra todas las cadenas en la colección con un prefijo dado

TRIES

TRIES ESTANDAR



CARACTERISTICAS

- Ninguna cadena es prefijo de otra.
- Cada nodo excepto la raíz tiene una etiqueta que representa un caracter de Σ .
- Los caminos de la raíz a las hojas representan cadenas del conjunto.

PROPIEDADES

- Altura = Longitud de la cadena + larga.
- Cada nodo interno tiene max. Σ hijos.
- Una hoja asociada a cada cadena.
- Tamaño del Trie = $n + 1$ (n = long total de las cadenas).

TIEMPO DE EJECUCIÓN

- Búsqueda de una cadena de longitud $m = O(m \cdot |\Sigma|)$
 - Mejorar la eficiencia:
 - Asignar caracteres a los hijos usando
 - Tabla de búsqueda secund.
 - Tabla Hash
- ↓
- Reduce el tiempo en cada nodo a $O(\log |\Sigma|)$ o $O(1)$ en promedio
- ↓
- Expectativa final = $O(m)$

COINCIDENCIA DE PALABRAS

- Se busca si un patrón dado coincide exactamente con una de las palabras.
- Difiere de la coincidencia de patrones porque el patrón debe coincidir con una palabra entera en vez de una subcadena.
- Cada palabra en el texto debe agregarse al trie para lograr esta coincidencia.

CONSTRUCCION DE UN TRIE ESTANDAR

- Algoritmo incremental para insertar cadenas una por una.
- Ninguna cadena en S es subprefijo de otra.
- Durante la inserción de una cadena X se traza la ruta asociada creando nodos adicionales según sea necesario.
- El tiempo de ejecución para insertar es igual al de búsqueda.
- La construcción completa de un Trie para el conjunto S requiere un tiempo esperado de $O(n)$, n = long de cadenas de S .

TRIES COMPRIMIDOS



CARACTERISTICAS

- Un trie comprimido que almacena una colección S de s cadenas de un alfabeto de tamaño d .
- Cada nodo interno de T tiene al menos 2 hijos o la mayoría de los hijos tiene d hijos.
- Tiene hojas
- El número de nodos de T es $O(s)$.

COMPRESIÓN

- Los nodos con un único hijo se fusionan con su hijo
 - ✓ Reduce el consumo de memoria
 - ✓ Mejora la eficiencia en almacenamiento
- Los nodos se etiquetan con cadenas que son subcadenas de cadenas de la colección

Nodo redundante

- 1: Es un nodo interno
- 2: Tiene exactamente 1 hijo
- 3: No es la raíz del Trie

USO COMO ESTRUCTURA DE INDICE AUXILIAR

- ✓ Reduce el espacio total del trie para el estandar $O(n)$ a $O(s)$ para el trie comprimido
- n de cadenas en S
- long de cadenas en S

✗ La búsqueda en un trie comprimido no es más rápida que en un estandar ya que todavía es necesario comparar cada caracter del patrón buscado con las etiquetas potencialmente multicaracteres

✗ Ineficiencias espaciales cuando un nodo tiene un solo hijo.