

Tries

Nicolás D'Ippolito^{1,2}, Ariel Bendersky^{1,2}

¹Instituto de Investigaciones en Ciencias de la Computación - ICC, CONICET, Argentina.

²Departamento de Computación, FCEyN, Universidad de Buenos Aires, Buenos Aires, Argentina.

Algoritmos y Estructuras de Datos II
Segundo cuatrimestre de 2019

(2) Cómo harían el caso de las búsquedas?

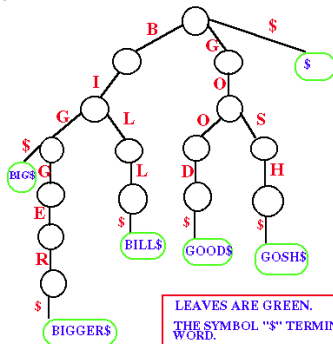
- El caso es el que uno tipeando y aparecen las palabras que coinciden...
- ¿Con un ABB? ¿Con un AVL? ¿Con una tabla de hash?
- ¿Qué complejidad tienen esas búsquedas?
- Hoy vamos a ver una estructura que sirve para esos escenarios, además de permitir otras varias aplicaciones.

(3) Suposiciones

- Esta estructura de datos requiere concentrarse en las claves.
- Vamos a suponer que las claves x pertenecen a un alfabeto Σ^* .
- Tenemos que pensar que no tienen todas $O(1)$, sino que su tienen una longitud $|x|$.
- Esa longitud puede ser:
 - La longitud propiamente dicha, en caso de strings.
 - $O(\log(x))$ si x es un entero.
- Además, definimos a $k = |\Sigma|$, como la cantidad de símbolos distintos del alfabeto Σ .
- Pensemos en árboles k -arios.

(4) Tries

Representemos {BIG, BIGGER, BILL, GOOD, GOSH}.

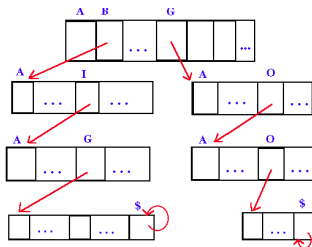


Debidos a Edward Fredkin, años '60. El nombre proviene de "retrieval".

(5) Análisis

- ¿Cuál es la complejidad de insertar clave x ? $O(|x|)$
- ¿Y de buscar? $O(|x|)$.
- Notar: es independiente de la cantidad de elementos almacenados.
- Algoritmos: muy simples.
- Almacenamiento: punto flojo.

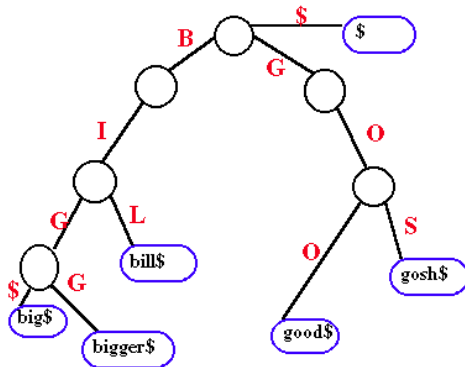
(6) Representación con arreglos y punteros



- Notar: puede ser muy ineficiente en términos de espacio.
- En especial cuando $|\Sigma|$ es grande y n es chico.
- ¿Podemos “comprimir” el arreglo? ¿Y qué pasa con la complejidad?
- En el peor caso se multiplica por k .

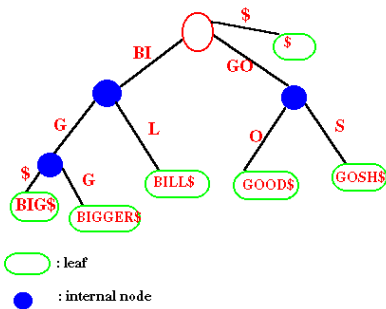
(7) Compactos

¿Qué pasa si compactamos las cadenas que sólo llevan a hojas?



(8) PATRICIA

- PATRICIA = *Practical Algorithm To Retrieve Information Coded In Alphanumeric*
- Debidos a D. R. Morrison.
- Ahora colapsamos todas las cadenas.
- Un eje puede representar más de un caracter.



(9) Vimos hoy

Vimos hoy:

- Tries: Árboles para recuperación de información.
- Mejoras colapsando ramas que sólo llevan a hojas.
- PATRICIA: Una mejora adicional a este tipo de árboles.

Lo que viene:

- Hashing.