

Tries

Son índices que se estructuran en función de la composición de caracteres de las claves de recuperación de registros de datos. El nombre deriva de la palabra en inglés “**retrieval**” (recuperación).

Para implementar tries en archivos hay básicamente dos posibilidades: usar registros de longitud fija, o usar registros de longitud variable. En el primer caso el registro es un arreglo de números de registro del mismo trie, con una posición para cada letra del alfabeto (27), más un puntero para el caracter de terminación, que indica el fin de una palabra y apunta al registro de datos. La raíz del trie es siempre el registro 0, y los números de registro nulos se representan con el 0.

Por ejemplo, para indexar con un trie de registros de longitud fija, registros de datos con claves Maria, Mary y Mario (con referencias r1, r2 y r3 respectivamente, que pueden ser números de registro, números de bloque o identificadores primarios):

```
0: 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1: 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 3 0 0 0 0 0 0 0 0 0
3: 0 0 0 0 0 0 0 0 0 4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 6 0 0
4: 5 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 7 0 0 0 0 0 0 0 0 0 0 0 0 0
5: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 r1
6: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 r2
7: 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 r3
```

Para registros de longitud variable, en vez de tener arreglos se tienen pares compuestos por un caracter y un número de nodo, y como los registros se deben organizar en bloques, cada registro debe tener como campo de control su número de identificación:

```
0(M, 1)
1(A, 2)
2(R, 3)
3(I, 4)(Y,6)
4(A, 5)(0, 7)
5(/n, r1)
6(/n, r2)
7(/n, r3)
```

Como a estos registros hay que organizarlos en bloques, se necesita un archivo de localización de registros con un registro por cada registro del trie, cuyo único campo es el número de bloque donde encontrar el registro del trie cuyo número coincida con la posición; por ejemplo si todos los registros anteriores estuvieran en el bloque 0, la tabla de localización quedaría 0 0 0 0 0 0 0 0. Dentro de los bloques los registros se buscan secuencialmente hasta localizarlos por su número.

También se puede fijar la cantidad de niveles del trie en un máximo determinado; supongamos que 3, entonces los nodos hoja van a contener cadenas de terminación de palabras junto con referencias a registros de datos:

```
0(M, 1)
1(A, 2)
2(3RIA, r1)(2RY, r2)(3RIO, r3)
```

Si alguna palabra tuviera menos de 3 caracteres, bastaría con poner el caracter de terminación en algún nivel anterior. Aquí conviene tener una política de asignación de bloques de almacenamiento de manera de no mezclar registros hoja del trie con registros internos en un mismo bloque.