

# Resumen práctica hashing

la dispersión de archivos es una técnica para generar una dirección base única para una clave dada. Convierte la clave en un número aleatorio que luego sirve para determinar dónde se almacena la clave.

usa una función de dispersión para mapear cada clave con una dirección física de almacenamiento

es usada cuando se requiere acceso rápido por clave

## TIPOS DE DISPERSIÓN

- 1) **direccionamiento estático:** el espacio disponible para dispersar los registros del archivo está fijado previamente
- 2) **direccionamiento dinámico:** el espacio disponible para dispersar aumenta o disminuye en función de las necesidades

## PARÁMETROS A CONSIDERAR

- capacidad de almacenamiento de cada dirección
- densidad de empaquetamiento: relación entre el espacio disponible para el archivo de datos y la cantidad de registros que integran el mismo.  $DE = \text{número de registros} / \text{espacio total}$
- función de hash: caja negra que a partir de una clave genera la dirección física donde debe almacenarse el registro
- método de tratamiento de desbordes: es la situación en la cual una clave carece de lugar en la dirección asignada por la función de dispersión
- colisión: situación en la que un registro es asignado, por función de dispersión, a una dirección que ya posee uno o más registros

aunque la función de dispersión sea eficiente y la densidad de empaquetamiento sea baja, a veces pueden ocurrir desbordes. Los cuales son tratados de diferentes maneras

### SATURACIÓN PROGRESIVA

cuando se completa una dirección de memoria se busca en las siguientes direcciones en secuencia, hasta encontrar una vacía

su ventaja es la **simplicidad**

- búsqueda: comienza en la dirección base y continúa buscando en localidades sucesivas hasta encontrar la clave buscada
- eliminación: no debe permitirse que el espacio liberado por la eliminación obstaculice las búsquedas posteriores, al mismo tiempo debe ser posible usar el espacio liberado para adiciones posteriores. Se marca el espacio liberado con ### (el espacio liberado no rompe la secuencia de búsquedas y el liberado está disponible y puede ser usado en adiciones posteriores) cuando se elimina clave de una cubeta lleva y en la próxima cubeta hay otra marca u otro dato

### SATURACIÓN PROGRESIVA ENCADENADA

funciona igual que la progresiva excepto que las claves sinónimos se enlazan por apuntadores

cada dirección base contiene un número que indica el lugar del siguiente registro con la misma dirección base. El siguiente registro contiene a la vez un puntero al siguiente registro con la misma dirección base y así sucesivamente

su ventaja es que solo se necesita acceder a los registros con claves que son sinónimos por lo tanto mejora el número de accesos promedio

su desventaja es que debe agregarse un campo de enlace a cada registro y requerirá mayor espacio de almacenamiento

*importante:* si la clave en la dirección base a insertar es intrusa, se le asigna nuevo lugar y a partir de la dirección base de la intrusa, se modifica el enlace a su nueva dirección luego se inserta la nueva clave en su dirección base

### SATUACIÓN PROGRESIVA CON ENCADENAMIENTO EN ÁREAS SEPARADAS

cuando se agrega un registro nuevo si hay lugar en la dirección base se almacena allí, sino se mueve al archivo de saturación donde se agrega a la lista enlazada que comienza en la dirección base

*importante:* si la baja es en el área principal, solamente escribo la cubeta sin la clave a eliminar, no se modifican los enlaces

### DISPERSIÓN DOBLE

Cuando sucede una colisión se aplica una segunda función de dispersión a la clave, para producir un N.º, el cual se suma a la dirección original tantas veces como sea necesario hasta encontrar una dirección vacía

- eliminación: se usan marcas ### para futuras búsquedas