

BÚSQUEDA POR TRANSFORMACIÓN

Tablas de dispersión – Organización

Hashing - asociación; y en sus comienzos aleatorización, randomizing

Métodos de búsqueda

- Una gran parte del tiempo estamos buscando algo.
 - ¿Cómo podemos minimizar el tiempo?
- Clasificación de los métodos
 - Interno
 - Externo
- Tipos de búsqueda
 - Secuencial :: $O(n)$:: I/E
 - Binaria :: $O(\log_2 n)$:: I
 - Por transformación :: $O(?)$:: I/E
 - Árboles de búsqueda B. (BST) :: $O(?)$:: I

¿Qué pasa si la cantidad de elementos se incrementa?

Una aproximación...

- Conjunto de elementos (claves) reducido
 - Ej: Trabajadores , legajos 1 .. 600

...
551
552
...
...
599
600

Legajo: 551
DNI: 12.345.678
Nombre: pepe
...
...

¿ Si la clave fuese otra ?
ej: DNI

¿ Si la clave no fuese numérica?

¿ Si el conjunto crece ?

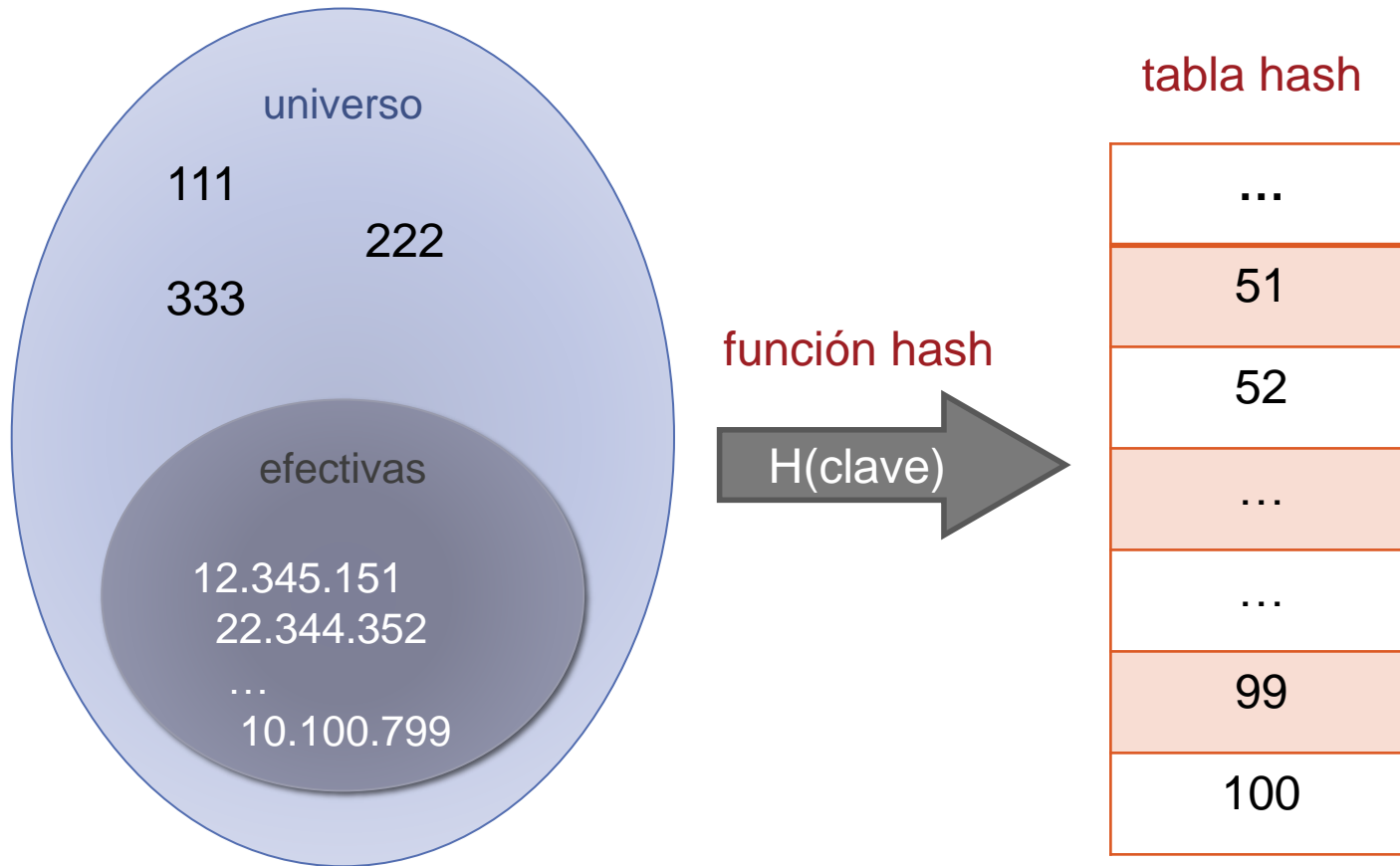
Transformación de clave - hash

- Función que convierte la clave en una dirección

$$\text{dirección} \leftarrow H(\text{clave})$$

- Operaciones
 - Inserción, Eliminación y Búsqueda
- La elección de la función debe mantener un equilibrio
 - Fácil de calcular
 - Distribución uniforme
- ¿ Es fácil encontrar la función ?
 - Hacer análisis de las principales características

Transformación de clave - hash



Transformación de clave - hash

- Funciones más utilizadas:

- Módulo – división

- $H(k) = (k \bmod \max) + 1$

- Cuadrado – central

- $H(k) = \text{dig_centrales}(k^2) + 1$

- Plegamiento

- $H(k) = \text{prim}(k) [+/*] \text{seg}(k) + 1$

- Truncamiento

- $H(k) = \text{eleccion_digitos}(k) + 1$

clave **alfanumérica**

- suma valores de representación
- método Horner combinado

¿Qué pasa si existe una misma dirección para dos o más claves?


Qué sucede si

$$(K_a \neq K_b) \ \& \ (H(K_a) = H(K_b))$$

La elección de la función no me garantiza que no exista.

Tratamiento de colisión

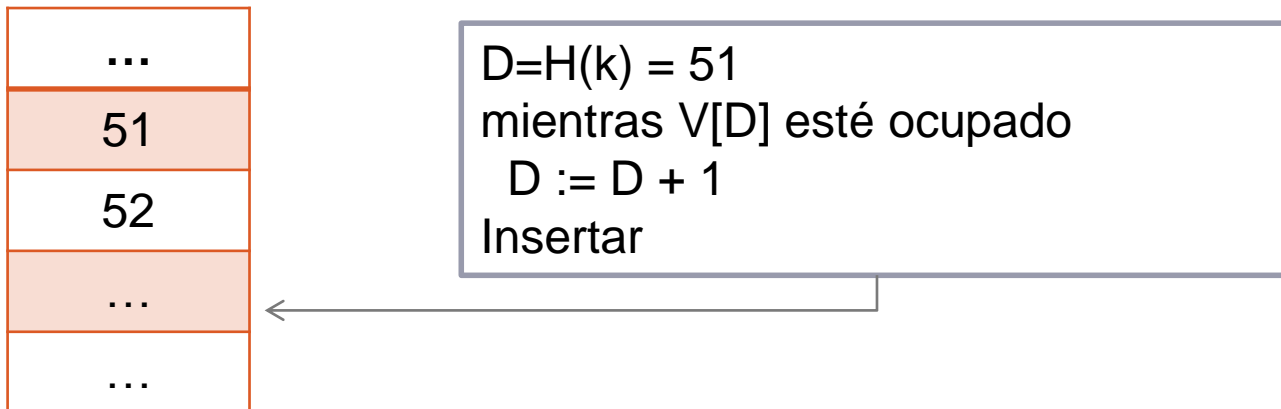
- La elección del método para tratar la colisión es tan importante como la elección de la función
- Métodos más utilizados
 - Reasignación
 - Prueba lineal
 - Prueba cuadrática
 - Re-hashing o doble dirección
 - Arreglos anidados
 - Encadenamiento



Hay que tener en cuenta el método al momento de buscar un elemento

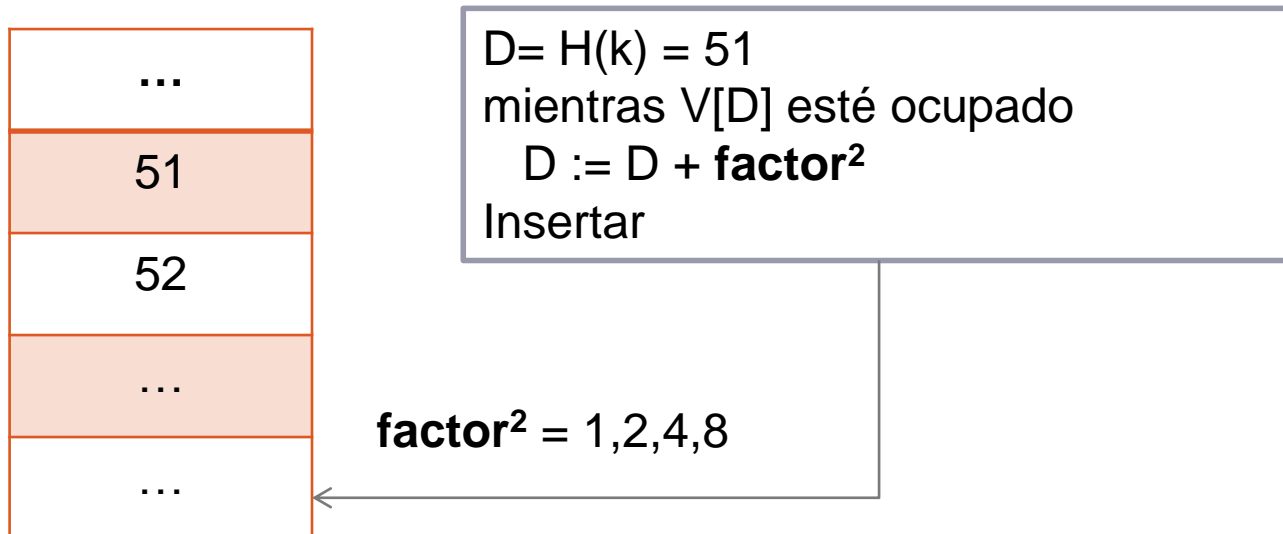
Tratamiento de colisión

- **Direcccionamiento Abierto – Hashing Cerrado – Dispersión cerrada.**
 - Cada espacio del arreglo contendrá un elemento o estará vacío. La estructura se trata de forma circular.
 - **Prueba lineal:** una vez detectada la colisión se busca el próximo espacio vacío.
 - Desventaja: fuerte agrupamiento alrededor de ciertas claves



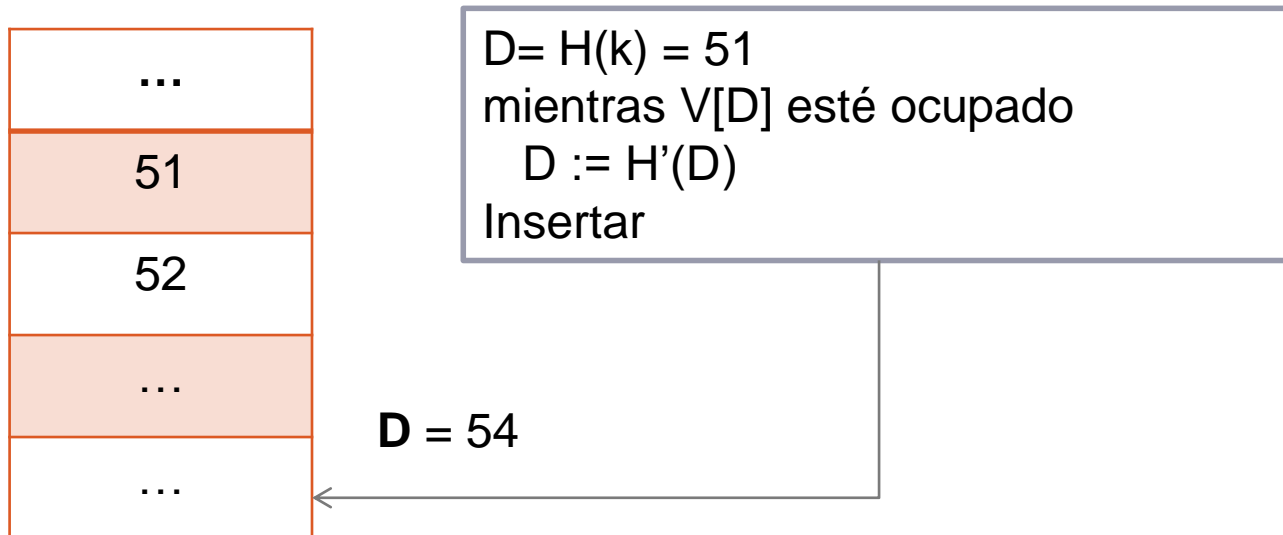
Tratamiento de colisión

- **Prueba cuadrática:** una vez detectada la colisión se busca la dirección más un factor que irá aumentando de forma cuadrada.
- Ventaja: Mejora la distribución



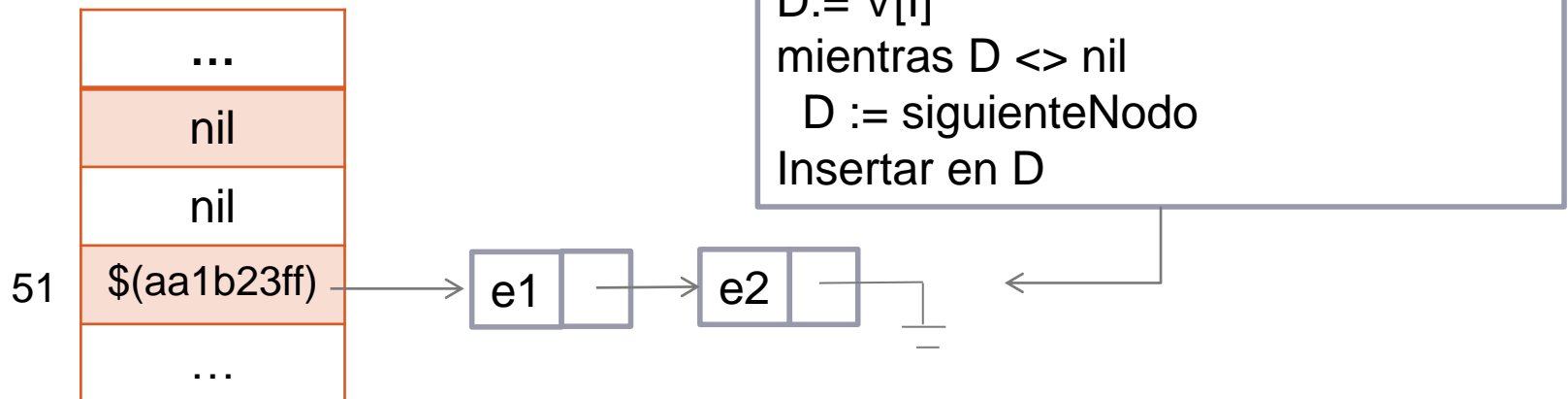
Tratamiento de colisión

- **Doble dirección – re-hashing:** una vez detectada la colisión se aplica nuevamente la función a la dirección obtenida. Si hay colisión la función puede ser diferente a la primera.



Tratamiento de colisión

- **Direcccionamiento Cerrado – Hashing Abierto – Dispersión abierta.**
 - Cada espacio del arreglo contendrá un puntero a **una lista enlazada**. En caso de colisión se inserta en la lista.
 - Desventaja: Mismo problemas q las listas. Espacio adicional.
 - El peor caso $O(n)$



Algunas conclusiones

- Permite aumentar la velocidad de búsqueda sin tener los elementos ordenados.
- Se pierde la noción de orden.
- Tiempo de búsqueda independiente del conjunto de elementos.
- En promedio el acceso es de $O(1)$.
- El tiempo de cálculo de la función se debe tener en cuenta.
- NOTA: Existen técnicas de expansión y disminución.

Anexo – Método Horner o Regla H.

- William George Horner, matemático, no fue el primero en encontrar el método pero lleva su nombre por su publicación.
- Aplica a la solución para un valor x conocido de un polinomio, si bien la solución clásica es sencilla el método reduce la complejidad algorítmica.
- Sea el grado de un polinomio n :
 - Método tradicional: $(n^2+n)/2$ multiplicaciones y n sumas.
 - Método Horner: n multiplicaciones y n sumas.

Anexo – Método Horner o Regla H.

$$4x^4 + 7x^3 + 3x^2 + 6x + 2$$

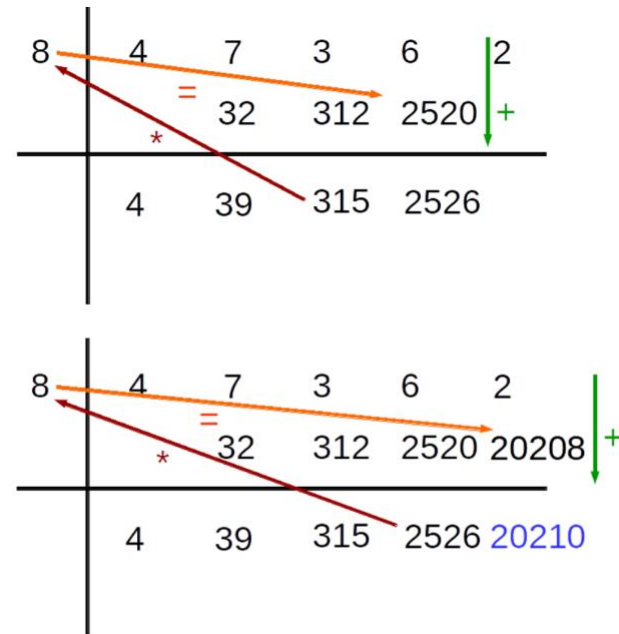
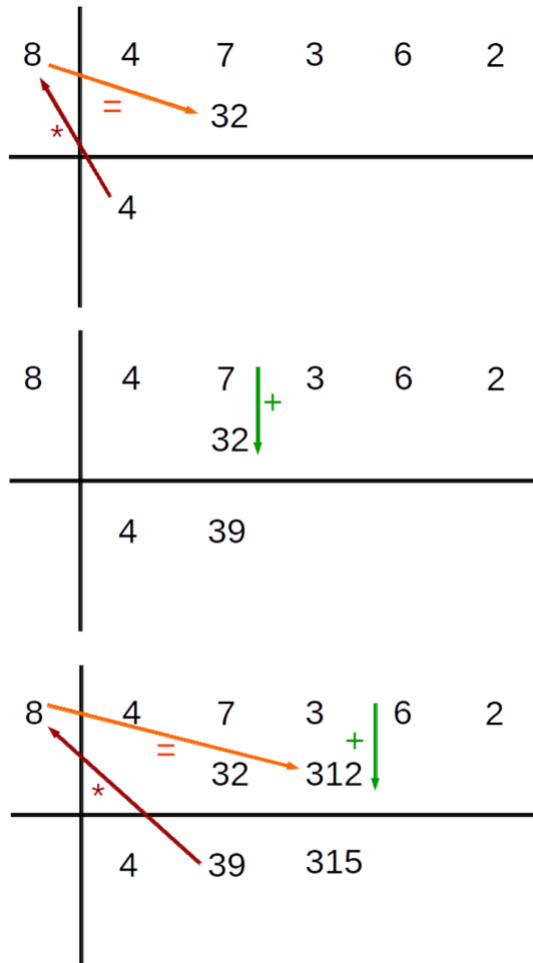
Si $x = 8 \rightarrow$

$$4(8)^4 + 7(8)^3 + 3(8)^2 + 6(8) + 2$$

$$4*(8*8*8*8) + 7*(8*8*8) + 3*(8*8) + 6*8 + 2$$

$$163854 + 3584 + 192 + 48 + 2 = 20210$$

Anexo – Método Horner o Regla H.



Anexo – Método ...en Pascal

```
function hashing(const clave:string):longint;  
var i:byte;  
begin  
    hashing := 0;  
    for i:= 1 to length(clave)do  
        hashing := (hashing *32 +  
                    ord(clave[i]))mod N;  
    end;
```

Anexo – Bibliografía Hashing

- Tratamiento de colisión en el mismo arreglo
 - Mary E. S. Loomis - direccionamiento abierto – hashing cerrado.
 - Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman – Dispersión cerrada.
- Tratamiento de colisión con listas simplemente encadenadas
 - Xavier F. Gutierrez – tablas de dispersión encadenadas abiertas – tablas de dispersión encadenadas indirectas.
 - Osvaldo Cairó, Silvia Guardati - Encadenamiento
- Abstracción y estructuras de datos en C++ - Carrillo y Valdivia