

Tablas de Hash

- . La dispersión es una técnica empleada para realizar inserciones, eliminaciones y búsquedas en un tiempo promedio constante
- . La estructura de datos ideal para la tabla de dispersión es simplemente una lista de tamaño fijo N que contiene las claves
- . Toda función de dispersión debe:
 - . Calcularse de forma sencilla O (1).
 - . Distribuir uniformemente las claves

Resolución de colisiones: dispersión abierta

- . La estrategia consiste en tener una lista de todos los elementos que se dispersan en el mismo valor
- . Para efectuar un insertar, recorremos la lista adecuada para revisar si el elemento ya está en la lista. Si el elemento resulta ser nuevo, se inserta al frente o al final de la lista
- . Básicamente pasas el valor por la función del hash y la insertas en el lugar correspondiente, en este caso las colisiones se resuelven fácilmente ya que disponemos de una lista de listas.

Valores de la función de dispersión:

$\text{hash}(\text{Ana}, 11) = 7$

$\text{hash}(\text{Luis}, 11) = 6$

$\text{hash}(\text{José}, 11) = 7$

Tabla después de insertar *Ana*:

0	1	2	3	4	5	6	7	8	9	10
[]	[]	[]	[]	[]	[]	[]	[Ana]	[]	[]	[]

Tabla después de insertar *Luis*:

0	1	2	3	4	5	6	7	8	9	10
[]	[]	[]	[]	[]	[]	[Luis]	[Ana]	[]	[]	[]

Tabla después de insertar *José*:

0	1	2	3	4	5	6	7	8	9	10
[]	[]	[]	[]	[]	[]	[Luis]	[Ana; José]	[]	[]	[]

Ejemplardo

Resolución de colisiones: dispersión cerrada

- . En un sistema de dispersión cerrada, si ocurre una colisión, se intenta buscar celdas alternativas hasta encontrar una vacía
- . Como todos los datos se guardan en la tabla, esta tiene que ser más grande para la dispersión cerrada que para la abierta

Dispersión cerrada con exploración lineal

- . Insertas normalmente, y si se produce una colisión hay que buscar secuencialmente en la lista (en forma circular) de uno en uno hasta encontrar una posición vacía.
- . Si la tabla es lo suficientemente grande, siempre se encontrará una celda vacía

Valores de la función de dispersión:

$\text{hash}(\text{Ana}, 11) = 7$

$\text{hash}(\text{Luis}, 11) = 6$

$\text{hash}(\text{José}, 11) = 7$

Tabla de dispersión después de insertar *Ana*:

0	1	2	3	4	5	6	7	8	9	10
							Ana			

Tabla de dispersión después de insertar *Luis*:

0	1	2	3	4	5	6	7	8	9	10
						Luis	Ana			

Tabla de dispersión después de insertar *José*:

0	1	2	3	4	5	6	7	8	9	10
						Luis	Ana	José		

Ejemplardo

Dispersión cerrada con exploración cuadrática

- Insertas normalmente, y si se produce una colisión hay que ir probando si encontramos celda vacía avanzando de i^2 siendo i la cantidad de colisiones
- La función de resolución de colisiones es cuadrática, por lo general:
 $f(i) = i^2$
- Con exploración lineal es malo llenar la tabla, porque se degrada el rendimiento, Para la exploración cuadrática, la situación es más drástica:
- Con mas de la mitad de la tabla ocupada, no hay garantías de encontrar

Ejemplardo

Valores de la función de dispersión:

hash(Ana,11) =7 hash(Luis,11)=6 hash(José,11)=7
hash(Olga,11)=7 hash(Rosa,11)=6 hash(Iván,11)=6

Tabla de dispersión después de insertar *Ana*:

0	1	2	3	4	5	6	7	8	9	10
							Ana			

Tabla de dispersión después de insertar *Luis*:

0	1	2	3	4	5	6	7	8	9	10
						Luis	Ana			

Tabla de dispersión después de insertar *José*: 1^2

0	1	2	3	4	5	6	7	8	9	10
						Luis	Ana	José		

Tabla de dispersión después de insertar *Olga*: 2^2

0	1	2	3	4	5	6	7	8	9	10
Olga						Luis	Ana	José		

Tabla de dispersión después de insertar *Rosa*: 2^2

0	1	2	3	4	5	6	7	8	9	10
Olga						Luis	Ana	José		Rosa

Tabla de dispersión después de insertar *Iván*: 3^2

0	1	2	3	4	5	6	7	8	9	10
Olga			Iván			Luis	Ana	José		Rosa

Dispersión cerrada con exploración doble

- Acá se van a utilizar dos funciones de dispersión: El hash original, y el hash2 que nos dará el tamaño de salto que vamos a pegar si es que se producen colisiones.
- Entonces, para la resolución de colisiones aplicamos una segunda función de dispersión, en general: $f(i) = i * h_2(x)$. Donde i es la cantidad de colisiones.
- Una función como $h_2(x) = R * (x \% R)$, con R un número primo menor que el tamaño de la tabla, funcionará bien.
- Hay que asegurarse de que el tamaño de la tabla sea primo.

Ejemplardo

Valores de la función de dispersión:

$$h_1(x, 11)$$

$$h_2(x, 11) = 5 - h_1(x, 11) \% 5$$

Ana	Luis	José	Olga	Rosa	Iván
7	6	7	7	6	6
3	4	3	3	4	4

Tabla de dispersión después de insertar *Ana*:

0	1	2	3	4	5	6	7	8	9	10
							Ana			

Tabla de dispersión después de insertar *Luis*:

0	1	2	3	4	5	6	7	8	9	10
						Luis	Ana			

Tabla de dispersión después de insertar *José*: $7 + (1 * 3) = 10$

$H1(i * h_2)$

0	1	2	3	4	5	6	7	8	9	10
						Luis	Ana			José

Tabla de dispersión después de insertar *Olga*:

$$7 + (2 * 3) = 13 \% 11 = 2$$

0	1	2	3	4	5	6	7	8	9	10
		Olga				Luis	Ana			José

Tabla de dispersión después de insertar *Rosa*:

$$6 + (2 * 4) = 14 \% 11 = 3$$

0	1	2	3	4	5	6	7	8	9	10
		Olga	Rosa			Luis	Ana			José

Tabla de dispersión después de insertar *Iván*:

$$6 + (4 * 4) = 22 \% 11 = 0$$

0	1	2	3	4	5	6	7	8	9	10
Iván		Olga	Rosa			Luis	Ana			José

Dispersión del cuco

- La inserción de una nueva clave en una tabla de dispersión puede empujar a otra clave para una ubicación diferente en la tabla.
- La idea básica es usar dos funciones de dispersión en lugar de sólo una. Esto proporciona dos posibles ubicaciones en la tabla para cada clave única y tiempos constantes en la búsqueda de valores.
- La implementación más común es la de dividir la tabla de dispersión en dos tablas más pequeñas de igual tamaño, y cada función dispersión proporciona un índice en una de estas dos tablas.
- Cuando se inserta una nueva clave, si esta operación no produce colisión se realiza sin inconvenientes.
- Si existen valores ocupando una de sus dos posibles ubicaciones, "se patea", es decir, desplazar, cualquier clave que podría residir en estas ubicaciones y se inserta esta clave desplazada en su lugar alternativo.
- El procedimiento de expulsión continua hasta que se encuentre un puesto disponible para todas las claves expulsadas, o se genere

Ejemplardo

Valores de las funciones de dispersión:

$$h_1(\text{key}) = \text{key} \% 11$$

$$h_2(\text{key}) = (\text{key}/11) \% 11$$

	20	50	53	75	100	67	105	3	36	39
$h_1(\text{key})$	9	6	9	9	1	1	6	3	3	6
$h_2(\text{key})$	1	4	4	6	9	6	9	0	3	3

Se comienza insertando el 20 en la primera tabla en la posible posición indicada por $h_1(20)$:

table[1]	-	-	-	-	-	-	-	-	20	-
table[2]	-	-	-	-	-	-	-	-	-	-

Sigue el 50:

table[1]	-	-	-	-	-	50	-	-	20	-
table[2]	-	-	-	-	-	-	-	-	-	-

Sigue el 53, pero su ubicación esta ocupada por el 20, entonces el 53 se ubica en la tabla 1 y el 20 en la tabla 2:

table[1]	-	-	-	-	-	50	-	-	53	-
table[2]	-	20	-	-	-	-	-	-	-	-

Valores de las funciones de dispersión:

$$h_1(\text{key}) = \text{key} \% 11$$

$$h_2(\text{key}) = (\text{key}/11) \% 11$$

	20	50	53	75	100	67	105	3	36	39
$h_1(\text{key})$	9	6	9	9	1	1	6	3	3	6
$h_2(\text{key})$	1	4	4	6	9	6	9	0	3	3

Sigue el 75, pero el 53 ocupa su lugar, Entonces el 75 se almacena en la tabla 1 y el 53 en la tabla 2.

table[1]	-	-	-	-	-	-	50	-	-	75	-
table[2]	-	20	-	-	53	-	-	-	-	-	-

Sigue el 100.

table[1]	-	100	-	-	-	-	50	-	-	75	-
table[2]	-	20	-	-	53	-	-	-	-	-	-

Sigue el 67, pero el 100 ocupa su lugar, Entonces el 67 se almacena en la tabla 1 y el 100 en la tabla 2.

table[1]	-	67	-	-	-	-	50	-	-	75	-
table[2]	-	20	-	-	53	-	-	-	-	100	-

Valores de las funciones de dispersión:

$$h_1(\text{key}) = \text{key} \% 11$$

$$h_2(\text{key}) = (\text{key}/11) \% 11$$

	20	50	53	75	100	67	105	3	36	39
$h_1(\text{key})$	9	6	9	9	1	1	6	3	3	6
$h_2(\text{key})$	1	4	4	6	9	6	9	0	3	3

Sigue el 105, pero el 50 ocupa su lugar, entonces el 105 se ubica en la tabla 1 y el 50 en la tabla 2. Ahora el 53 ha sido desplazado a la tabla 1 en la posición 9 y el 75 a la tabla 2 en la posición 6.

table[1]	-	67	-	-	-	-	105	-	-	53	-
table[2]	-	20	-	-	50	-	75	-	-	100	-

Sigue el 3.

table[1]	-	67	-	3	-	-	105	-	-	53	-
table[2]	-	20	-	-	50	-	75	-	-	100	-

Sigue el 36. Quedando el 36 en la posición 3 de la tabla 1 y el 3 en la posición 0 de la tabla 2.

table[1]	-	67	-	36	-	-	105	-	-	53	-
table[2]	3	20	-	-	50	-	75	-	-	100	-

Valores de las funciones de dispersión:

$$h_1(\text{key}) = \text{key} \% 11$$

$$h_2(\text{key}) = (\text{key}/11) \% 11$$

	20	50	53	75	100	67	105	3	36	39
$h_1(\text{key})$	9	6	9	9	1	1	6	3	3	6
$h_2(\text{key})$	1	4	4	6	9	6	9	0	3	3

Finalmente se inserta el 39. Quedando el 39 en la tabla 1, el 105 en la tabla 2, el 100 en la tabla 1, el 67 en la tabla 2, el 75 en la tabla 1, el 53 en la tabla 2, el 50 en la tabla 1 y el 39 en la tabla 2.

table[1]	-	100	-	36	-	-	50	-	-	75	-
table[2]	3	20	-	39	53	-	67	-	-	105	-

Practica 4

Ejercicio 1

Dados los enteros 5, 20, 3, 1000, 45, 27, 25, una tabla hash de tamaño 11 y una estrategia de resolución de colisiones de dispersión abierta, con la siguiente función de dispersión que retorna una posible posición para el elemento "x":

```
public int h(int x){  
    return x%11;  
}
```

Graficar el estado de la tabla antes y después de insertar los elementos.

Función Hash:

$5 \% 11 = 5$ $45 \% 11 = 1$
 $20 \% 11 = 9$ $27 \% 11 = 5$
 $3 \% 11 = 3$ $25 \% 11 = 3$
 $1000 \% 11 = 10$

	45		3,25		5,27				20	1000
0	1	2	3	4	5	6	7	8	9	10

Ejercicio 2

Dados los enteros 5, 20, 3, 1000, 45, 27, 38, una tabla hash de tamaño 11 y una estrategia de resolución de colisiones de dispersión cerrada, con la siguiente función de dispersión que retorna una posible posición para el elemento "x":

```
public int h(int x){  
    return x%11;  
}
```

Graficar el estado de la tabla antes y después de insertar los elementos, bajo las siguientes resoluciones de colisiones:

- a. Resolución de colisiones lineal: $f(i) = i$ → Función de dispersión: $h_i(x) = (h(x) + f(i)) \% 11$
- b. Resolución de colisiones cuadrática: $f(i) = i^2$ → Función de dispersión: $h_i(x) = (h(x) + f(i)) \% 11$

Función Hash:

5 % 11 = 5

45 % 11 = 1

20 % 11 = 9

27 % 11 = 5

3 % 11 = 3

38 % 11 = 5

1000 % 11 = 10

Resolución de colisiones lineal

	45		3		5	27	38		20	1000
0	1	2	3	4	5	6	7	8	9	10

Función Hash:

5 % 11 = 5

45 % 11 = 1

20 % 11 = 9

27 % 11 = 5^{1²}

3 % 11 = 3

38 % 11 = 5^{5²}

1000 % 11 = 10

Resolución de colisiones cuadrática

	45		3		5	27		38	20	1000
0	1	2	3	4	5	6	7	8	9	10

Ejercicio 3

Dados los enteros 5, 20, 3, 1000, 45, 27, 25, 67, 105, 3, 36, 39 dos tablas de hash de tamaño 11, y una estrategia de resolución de colisiones de dispersión del cuco, con las siguientes funciones de dispersión que retornan posibles posiciones para el elemento "x":

- $h1(key) = key \% 11$
- $h2(key) = (key/11) \% 11$

Graficar el estado de cada tabla antes y después de insertar los elementos.

Resolución de colisiones del Cuco

Función Hash 1:

$5 \% 11 = 5$ $25 \% 11 = 3$
 $20 \% 11 = 9$ $67 \% 11 = 1$
 $3 \% 11 = 3$ $105 \% 11 = 6$
 $1000 \% 11 = 10$ $36 \% 11 = 3$
 $45 \% 11 = 1$ $39 \% 11 = 6$
 $27 \% 11 = 5$

Función Hash 2:

$5/11 \% 11 = 0$ $25/11 \% 11 = 2$
 $20/11 \% 11 = 1$ $67/11 \% 11 = 6$
 $3/11 \% 11 = 0$ $105/11 \% 11 = 9$
 $1000/11 \% 11 = 2$ $36/11 \% 11 = 3$
 $45/11 \% 11 = 4$ $39/11 \% 11 = 3$
 $27/11 \% 11 = 2$

H1

3	67		36		27	39			20	1000
0	1	2	3	4	5	6	7	8	9	10

H2

5		25		45					105	
0	1	2	3	4	5	6	7	8	9	10

The End!!!