

 Centro Asociado de Gijón	Titulación	Ing. Técnica Informática	Página 1 de 3
	Asignatura	Estructura de Datos y Algoritmos	
	Tema	Transformaciones de Llaves - Wirth (285)	
	Examen	Febrero 1995; 1ª semana	
	Autor	César Menéndez Fernández	

Explicar detalladamente en qué consisten las tablas de transformación de llaves (hashing), cuáles son sus características y las ideas fundamentales necesarias para elegir una función de transformación. Explicar qué es una colisión y cómo se realiza el manejo de colisiones en general. Detallar la exploración lineal y cuadrática.

El hashing consiste en asociar la llave de un registro con el número del registro donde se va a grabar este dato, para lo cual el contenido del campo llave se somete a un algoritmo (función de transformación) que proporciona el número de registro que le corresponde. Para calcular los índices se utiliza una función aleatoria que utilice pocas operaciones aritméticas básicas. Una buena función de transformación debe distribuir las llaves lo mas uniformemente posible dentro de la gama de valores índice. Una asignación habitual es

$$H(k) = \text{ORD}(k) \text{ MOD } N$$

Donde ORD(k) es el ordinal de la llave k y N el número de elementos del arreglo. Esta función tiene la propiedad de que los valores de las llaves están distribuidas uniformemente en el intervalo índice, y constituye la base de la mayor parte de las transformaciones de llaves. Si $N < \text{ORD}(k)$, esto es, la cardinalidad de k es mayor que el número de elementos se producen colisiones. Cuando las llaves son secuencias de letras, hay que escoger como N un número primo para buscar la distribución uniforme de índices.

Ejemplo:

Supongamos que se tienen conjuntos de datos, valores enteros, formados siempre por menos de 8 elementos, y que se dispone de un vector de 8 componentes para almacenarlos. Para cada valor k se utiliza la función de transformación $H(k)=k \text{ MOD } 8$

Datos Llave (N=6)

1	2	5	8	10	18	21
---	---	---	---	----	----	----

Hashing

↓	↓	↓	↓	↓	↓	↓
---	---	---	---	---	---	---

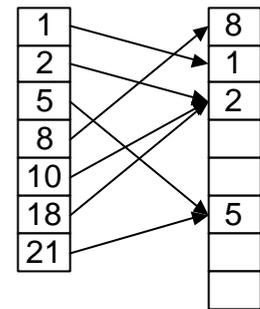
Orden de los datos en el arreglo final

1	2	5	0	2	2	5
---	---	---	---	---	---	---

Una colisión se produce cuando una llave, que no sea la buscada, se encuentra en la localización identificada. Así, en el ejemplo anterior, los datos 2, 10 y 18 tienen el mismo índice y no se pueden almacenar todos en la misma posición del vector. Lo mismo ocurre con el 5 y el 21. Cuando la función de transformación para un dato k indica una posición ya ocupada, se produce la colisión, y es necesario que dicho dato se cambie de posición para no solapar la información

del primero. Es decir, las colisiones se producen cuando la transformación inicial, que idealmente debía de ser uno a uno, es muchos a uno.

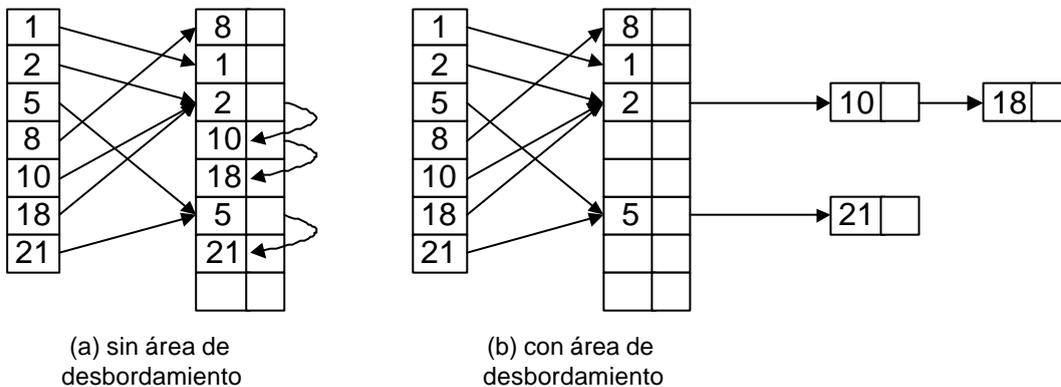
El manejo de colisiones se realiza generando índices alternos, habitualmente realizando un segundo sondeo y generando un nuevo índice a partir de la llave dada. Básicamente existen dos métodos para el manejo de colisiones:



Transformación de llaves

- Encadenamiento directo: Este método consiste en ligar todos los elementos con el índice primario idéntico $H(k)$ en una lista. Si los elementos de la lista no están en la tabla primaria, se da el nombre de área de desbordamiento al almacenamiento donde están asignados.

Encadenamiento Directo



- Dirección abierta: Este método consiste en observar las otras entradas en la misma tabla hasta encontrar el elemento o una posición vacía, en este caso se supondrá que la llave no se encuentra en la tabla.

El algoritmo de búsqueda en tablas, para el caso de *dirección abierta*, se escribirse como:

```

h:=H(k);i:=0;
REPEAT
  IF T[h].key = k THEN elemento encontrado
  ELSEIF T[h].key= vacia THEN el elemento no está en la tabla
  ELSE (* colisión *)
    i:=i+1;h:=H(k)+G(i)
  END
UNTIL encontrado ó no estaba en la tabla (ó tabla llena)
  
```

La selección de la función $G(i)$ es fundamental a la hora de realizar la exploración. Para aplicar el alguno de los métodos anteriores hay dos tipos de exploración:

Exploración lineal: consiste en ensayar la siguiente localización (suponiendo que la tabla es circular) hasta encontrar el elemento con la llave especificada o una localización vacía:

```

h0 = H(k)
hi = (h0 + i) MOD N , i = 1...N-1
  
```

Este método deja muy juntos los índices.

Exploración cuadrática: Evita la agrupación primaria que puede dar en la exploración lineal. No realiza la búsqueda en todas las tablas:

$$\begin{array}{l} h_0 = H(k) \\ h_i = (h_0 + i^2) \text{ MOD } N, \quad i > 0 \end{array}$$

Da saltos más grandes que la exploración lineal y realiza una mejor distribución.