

Archivos con Dispersión Extensible

PROBLEMA TIPO RESUELTO

Operaciones con registros representados por sus claves y los signos + para alta, y - para bajas, en un archivo directo con dispersión extensible con bloques o cubetas con capacidad para tres registros:

+123, +915, +629, +411, +200, +863, -629, +408, +34, +510, -863, +775

Resolución

La tabla se representa como una secuencia de números de bloque (se almacena como un archivo binario de números enteros sin signo), cuyas posiciones relativas (números de registro de la tabla), implícitas en la representación, comienzan en 0.

Los bloques se representan con su número seguido del carácter de dos puntos verticales como etiqueta (el número de bloque no es parte de su contenido), y a continuación el contenido, comenzando con los bits de dispersión del bloque entre paréntesis, y luego con las claves representantes de los registros contenidos en el bloque. Para implementar un archivo además es necesario otro campo de control para indicar la cantidad de registros cargados en el bloque, y si los registros fueran de longitud variable, también se requeriría de un indicador de bytes disponibles para almacenar más registros.

La función de dispersión depende del tamaño de la tabla y es $h(k)=k\%t$, donde h es la función de hash o dispersión, k la clave del registro a dispersar, y t el tamaño de la tabla. Los bits de dispersión de la tabla son $\log_2(t)$, que equivalen a la cantidad de bits menos significativos de la clave representada en binario que dan como resultado el módulo t de la clave. Por ejemplo, para un tamaño de tabla 8 y una clave 2197, $h(2197)=2197\%8=5$. Si se representa 2197 en binario, se observa que sus $\log_2(8)=3$ últimos bits equivalen al decimal 5: 100010010101.

+123, +915, +629, +411

Tabla: 0

Archivo:

0: (0) 123, 915, 629

Los bits de dispersión del bloque 0 son 0, porque todos los registros que contienen llegaron a él aplicando la función de dispersión $h(k)=k\%2^0$, es decir, considerando el módulo 1 de la clave.

El registro con clave 411 no cabe en el bloque que consigna la posición $411 \% 1=0$ de la tabla: 0

Se duplica la tabla y en la posición 0 se consigna un nuevo bloque: 1

Tabla: 1 0

Se incrementa en uno la cantidad de bits de dispersión del bloque 0 y se pone este último valor al bloque 1, y se redispersan los registros del bloque 0 con clave módulo 2:

Archivo:

0: (1) 123, 915, 629

1: (1)

Se reintentó agregar el registro con clave 411 y vuelve a desbordar el bloque 0. Se duplica la tabla y en la posición 1 se consigna un nuevo bloque: 2

Tabla: 1 2 1 0

Se incrementa en uno la cantidad de bits de dispersión del bloque 0 y se pone este último valor también al bloque 2, y se redispersan los registros del bloque 0 con clave módulo 4:

Archivo:

0: (2) 123, 915, 411

1: (1)

2: (2) 629

+200, +863

Tabla: 1 2 1 0

Archivo:

0: (2) 123, 915, 411

1: (1) 200

2: (2) 629

El registro con clave 863 no cabe en el bloque que consigna la posición $863 \% 4 = 3$ de la tabla: 0

Se duplica la tabla y en la posición 3 se consigna un nuevo bloque: 3

Tabla: 1 2 1 3 1 2 1 0

Se incrementa en 1 la cantidad de bits de dispersión del bloque 0 y se pone este último valor como cantidad de bits de dispersión del bloque 3, y se redispersan los registros del bloque 0 con clave módulo 8:

Archivo:

0: (3)

1: (1) 200

2: (2) 629

3: (3) 123, 915, 411

Se reintentará agregar el registro con clave 863: $863 \% 8 = 7$ --> va en el bloque 0:

Archivo:

0: (3) 863

1: (1) 200

2: (2) 629

3: (3) 123, 915, 411

-629

$629 \% 8 = 5$ --> se busca en el bloque 2, se borra, y éste queda vacío. El bloque 2 tiene 2 bits de dispersión, entonces a partir de la posición donde encontré consignado al bloque 2, la 5:

Posibilidad 1: me muevo $2^{2-1} = 2$ posiciones circularmente en la tabla, hacia arriba **o** hacia abajo, indistintamente (lo hago hacia arriba): $5 + 2 = 7$; en la posición 7 está consignado el bloque 0; la cantidad de bits de dispersión del bloque 0 es **distinta** que la del bloque 2, por lo que **NO PUEDO LIBERAR EL BLOQUE**.

Posibilidad 2: me muevo $2^{2-1} = 2$ posiciones circularmente en la tabla, hacia arriba **y** hacia abajo, y si en esas posiciones no está consignado el mismo bloque **NO PUEDO LIBERAR EL BLOQUE 2** (en este caso las posiciones 7 y 3 de la tabla consignan a los bloques 0 y 3 respectivamente, por lo que **NO PUEDO LIBERAR EL BLOQUE 2**).

Tabla: 1 2 1 3 1 2 1 0

Archivo:

0: (3) 863

1: (1) 200

2: (2)

3: (3) 123, 915, 411

+408, +34, +510:

Tabla: 1 2 1 3 1 2 1 0

Archivo:

0: (3) 863

1: (1) 200, 408, 34

2: (2)

3: (3) 123, 915, 411

El registro con clave 510 no cabe en el bloque consignado en la posición $510 \% 8 = 6$ de la tabla: 1

COMO LOS BITS DE DISPERSION DEL BLOQUE 1 SON MENORES QUE LOS DE LA TABLA (EL LOGARITMO BASE 2 DE SU TAMAÑO), NO SE DUPLICA LA TABLA: se suma 1 a los bits de dispersión del bloque 1 y se pone este mismo valor en los bits de dispersión a un bloque nuevo, el 4; se recorre la tabla circularmente desde la posición 6 a saltos de longitud equivalente a una potencia de 2 con exponente igual a los bits de dispersión del bloque nuevo ($2^2=4$), registrando este nuevo número de bloque en todas las posiciones, y se redispersa el contenido del bloque 1

Tabla: 1 2 4 3 1 2 4 0

Archivo:

0: (3) 863

1: (2) 200, 408

2: (2)

3: (3) 123, 915, 411

4: (2) 34

Se reintenta el alta del registro con clave 510:

Tabla: 1 2 4 3 1 2 4 0

Archivo:

0: (3) 863

1: (2) 200, 408

2: (2)

3: (3) 123, 915, 411

4: (2) 34, 510

-863

$863 \% 8 = 7$; el bloque 0, consignado en la posición 7, queda vacío. Sus bits de dispersión son 3, así que moviéndome circularmente en la tabla $2^{3-1}=4$ posiciones hacia adelante y hacia atrás encuentro en ambas posiciones (en realidad, la misma) consignado el mismo bloque: el 3; entonces PUEDO LIBERAR EL BLOQUE 0 (en la posición 7 consigno al bloque 3, recorro la tabla desde allí y a saltos circulares de longitud determinada por los bits de dispersión del bloque 3 consignando al bloque 3 (en un salto vuelvo a la misma posición, por lo que no hay que hacer ninguna sustitución más), resto 1 a la cantidad de bits de dispersión del bloque 3, y comparo las dos mitades de la tabla truncándola a la mitad si fueran iguales):

Tabla: 1 2 4 3 (1 2 4 3) <-- la segunda mitad de la tabla se trunca

Archivo:

0: (3) 863

1: (2) 200, 408

2: (2)

3: (2) 123, 915, 411

4: (2) 34, 510

Bloques libres: 0

+775

$775 \% 4=3$; en la posición 3 de la tabla está consignado el bloque 3, que no tiene más capacidad; como los bits de dispersión del bloque 3 (2) son iguales a los de la tabla ($\log_2 4=2$), se duplica la tabla, y en la posición 3 se consigna a un nuevo bloque (en este caso el 0, que se reutiliza); se suma 1 a los bits de dispersión del bloque 3, se pone este mismo valor como bits de dispersión del bloque 0, y se redispersan los registros del bloque 3:

Tabla: 1 2 4 0 1 2 4 3

Archivo:

0: (3) 123, 915, 411

1: (2) 200, 408

2: (2)

3: (3)

4: (2) 34, 510

Se reintentó el alta del registro con clave 775: $775 \% 8=7$: al bloque 3:

Tabla: 1 2 4 0 1 2 4 3

Archivo:

0: (3) 123, 915, 411

1: (2) 200, 408

2: (2)

3: (3) 775

4: (2) 34, 510