

CẤU TRÚC DỮ LIỆU & GIẢI THUẬT

HASH TABLE

Giảng viên: ThS Nguyễn Bích Vân

Giới thiệu

Đặt vấn đề

- Một tập hợp gồm nhiều phần tử và mỗi phần tử có 1 thuộc tính khóa (key) để xác định chúng.
- Cần tìm một cấu trúc phù hợp để lưu trữ các phần tử này sao cho chi phí thời gian tìm kiếm là tốt nhất → Bảng băm (Hash table)

Bảng băm

Bảng băm (hash table) là một cấu trúc dữ liệu có các đặc điểm sau:

- Có kích thước m xác định.
- Cho phép truy xuất ngẫu nhiên từng phần tử theo giá trị khóa.
- Độ phức tạp thời gian có thể đạt $O(1)$ khi truy xuất phần tử.

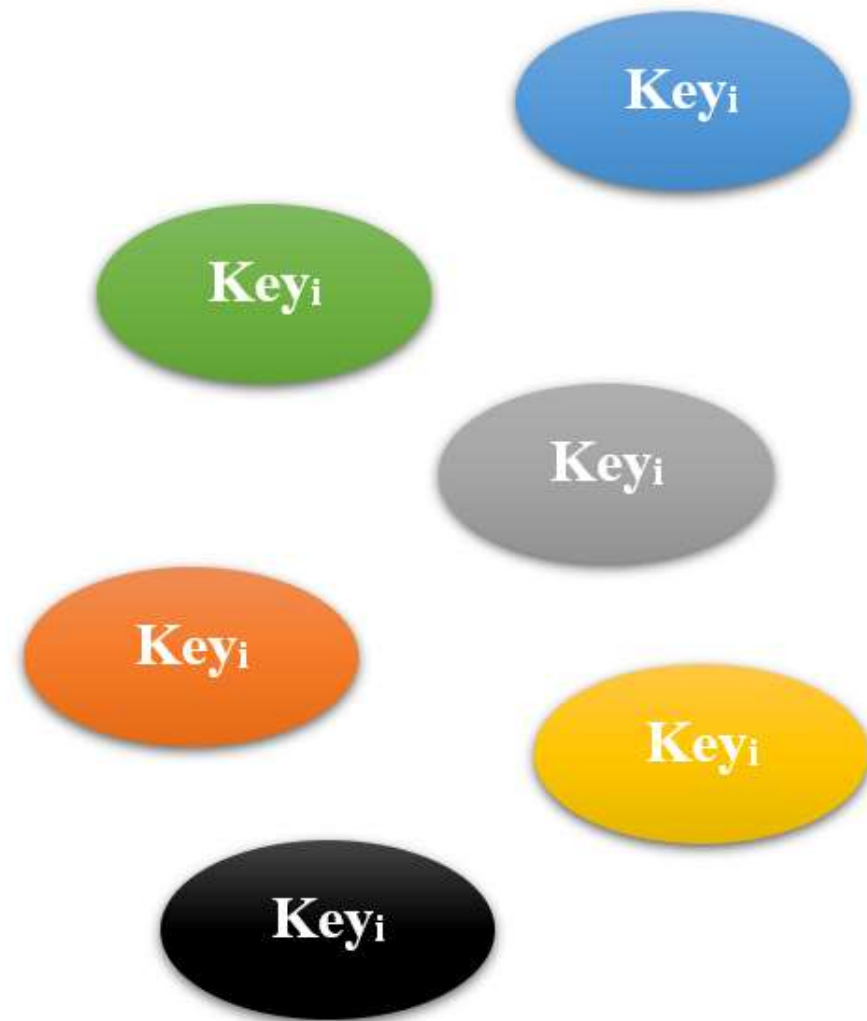
0
1
2
...
...
...
M-1

Một bảng băm tốt cần phải có hàm băm tốt.

Quá trình ánh xạ khóa vào bảng băm được thực hiện thông qua hàm băm

Bảng băm – Hash table

0	Key _i
1	Key _i
2	Key _i
3	Key _i
4	Key _i
...	...
...	...
...	...
...	...
...	...
M-1	Key _i



Hàm băm

Hàm băm $h(k)$ là hàm thực hiện việc ánh xạ khóa k nào đó vào trong bảng băm.

Một hàm băm tốt thỏa mãn các tiêu chí sau:

- Tốc độ tính toán nhanh
- Các khóa được phân bố đều trong bảng
- Ít xảy ra **đụng độ** \rightarrow xử lý **đụng độ**

0
1
2
...
...
...
M-1

Đụng độ (collision) là trường hợp hai giá trị khoá được ánh xạ vào cùng một giá trị chỉ số mảng.

Hàm băm

Đụng độ (collision) là trường hợp hai giá trị khoá được ánh xạ vào cùng một giá trị chỉ số mảng.

Hàm băm $h(k) = \text{key} \% M$ (với $M = 10$ là kích thước bảng băm)

Tập key = {10, 42, 15, 109, 63, 74, 28, 30}

0	10
1	
2	42
3	63
4	74
5	15
6	
7	
8	28
9	109

$$h(10) = 10 \% 10 = 0$$

$$h(42) = 42 \% 10 = 2$$

$$h(15) = 15 \% 10 = 5$$

$$h(109) = 109 \% 10 = 9$$

$$h(63) = 63 \% 10 = 3$$

$$h(74) = 74 \% 10 = 4$$

$$h(28) = 28 \% 10 = 8$$

$$h(30) = 30 \% 10 = 0 \rightarrow \text{đụng độ}$$

Hệ số tải (load factor)

của bảng băm T có kích thước m khi lưu trữ n phần tử là $\lambda = n/m$

VD: $n=8$ và $m=10$

$$\rightarrow \lambda = n/m = 0.8$$

Các dạng hàm băm

- Hàm băm dùng phép chia:

$$h(k) = k \% m$$

- Hàm băm dùng phép nhân:

$$h(k) = M * (k * A \% 1)$$

- Hàm băm phổ quát:

$$h(k) = \{h_{a,b}(k) = ((a * k + b) \% p) \% m\}$$

GIẢI QUYẾT ĐỤNG ĐỘ

Phương pháp nối kết

Nối kết trực tiếp

Nối kết hợp nhất

Phương pháp địa chỉ mở

Dò tuyến tính

Dò bậc 2

Băm kép

Phương pháp nối kết

Nối kết trực tiếp - Direct Changing

- Hàm băm: $H(\text{key}) = \text{key} \% M$
 - Giải quyết xung đột: gom thành 1 dslk
 - Ví dụ: Thêm các key 15, 23, 78, 25, 7, 56, 95, 43, 8, 13 vào bảng băm. Giải quyết xung đột bằng các phương pháp nối kết trực tiếp
- Cho kích thước bảng băm $M = 10$

Nối kết trực tiếp - Direct Chaining

Key = {15, 23, 78, 25, 7, 56, 95, 43, 8, 13} M=10

$$H(\text{key}) = \text{key} \% M$$

$$\text{Key} = 15 : H(15) = 15 \% 10 = 5$$

$$\text{Key} = 23 : H(23) = 23 \% 10 = 3$$

$$\text{Key} = 78 : H(78) = 78 \% 10 = 8$$

$$\text{Key} = 25 : H(25) = 25 \% 10 = 5 \text{ (đụng độ)}$$

$$\text{Key} = 7 : H(7) = 7 \% 10 = 7$$

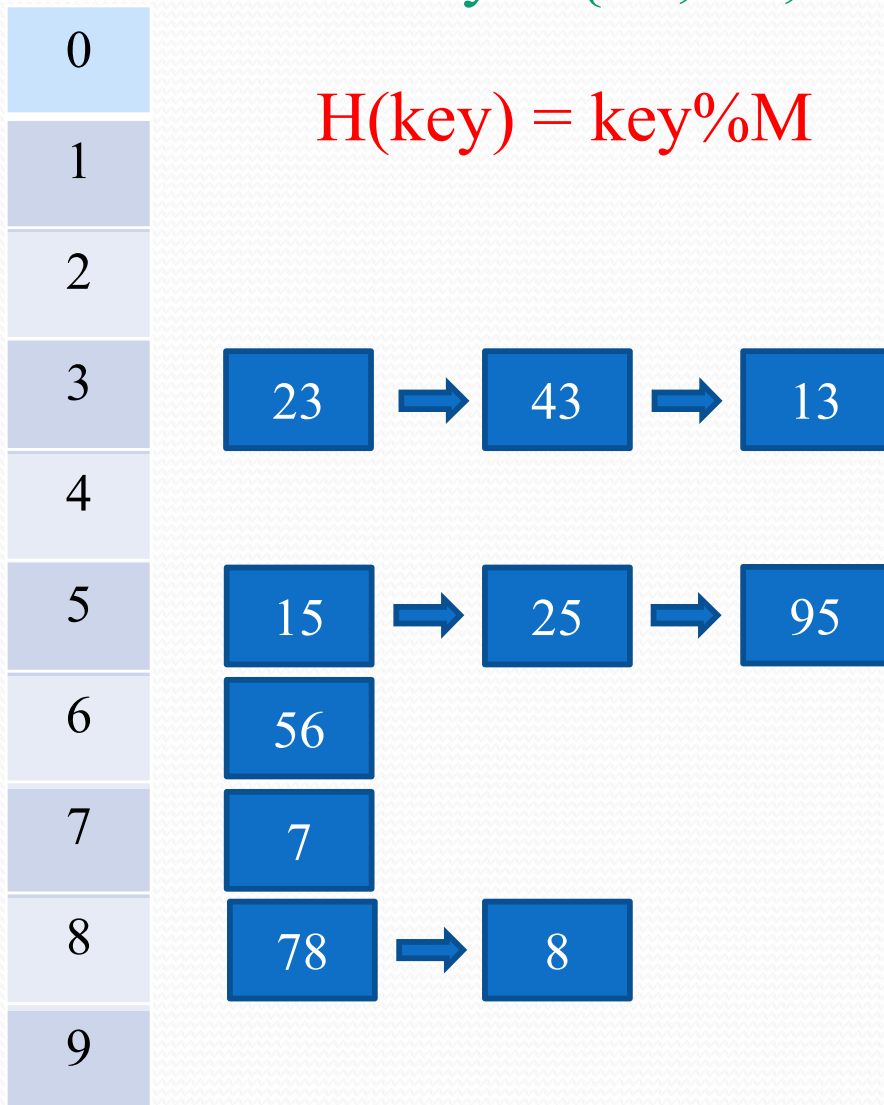
$$\text{Key} = 56 : H(56) = 56 \% 10 = 6$$

$$\text{Key} = 95 : H(95) = 95 \% 10 = 5 \text{ (đụng độ)}$$

$$\text{Key} = 43 : H(43) = 43 \% 10 = 3 \text{ (đụng độ)}$$

$$\text{Key} = 8 : H(8) = 8 \% 10 = 8 \text{ (đụng độ)}$$

$$\text{Key} = 13 : H(13) = 13 \% 10 = 3 \text{ (đụng độ)}$$



Phương pháp nối kết

Nối kết hợp nhất - Coalesced Chaining

- Hàm băm: $H(\text{key}) = \text{key} \% M$
 - Giải quyết xung đột: Nút trống phía cuối mảng
 - Ví dụ: Thêm các key 15, 23, 78, 25, 7, 56, 95, 43, 8, 13 vào bảng băm. Giải quyết xung đột bằng các phương pháp nối kết hợp nhất
- Cho kích thước bảng băm $M = 10$

Nối kết hợp nhất - Coalesced Chaining

Key = {15, 23, 78, 25, 7, 56, 95, 43, 8, 13} M=10

	Key	Next
0	NULL	-1
1	NULL	-1
2	NULL	-1
3	23	-1
4	NULL	-1
5	15	9
6	NULL	-1
7	NULL	-1
8	78	-1
9	25	-1

Key = 15 : $H(15) = 15 \% 10 = 5$

Key = 23 : $H(23) = 23 \% 10 = 3$

Key = 78 : $H(78) = 78 \% 10 = 8$

Key = 25 : $H(25) = 25 \% 10 = 5$

Nối kết hợp nhất - Coalesced Chaining

Key = {15, 23, 78, 25, 7, 56, 95, 43, 8, 13} M=10

	Key	Next
0	NULL	-1
1	NULL	-1
2	NULL	-1
3	23	-1
4	95	-1
5	15	9
6	56	-1
7	7	-1
8	78	1
9	25	4

Key = 15 : $H(15) = 15 \% 10 = 5$

Key = 23 : $H(23) = 23 \% 10 = 3$

Key = 78 : $H(78) = 78 \% 10 = 8$

Key = 25 : $H(25) = 25 \% 10 = 5$

Key = 7 : $H(7) = 7 \% 10 = 7$

Key = 56 : $H(56) = 56 \% 10 = 6$

Key = 95 : $H(95) = 95 \% 10 = 5$

Nối kết hợp nhất - Coalesced Changing

Key = {15, 23, 78, 25, 7, 56, 95, 43, 8, 13} M=10

	Key	Next
0	13	-1
1	8	-1
2	43	0
3	23	2
4	95	-1
5	15	9
6	56	-1
7	7	-1
8	78	1
9	25	4

Key = 15 : $H(15) = 15 \% 10 = 5$

Key = 23 : $H(23) = 23 \% 10 = 3$

Key = 78 : $H(78) = 78 \% 10 = 8$

Key = 25 : $H(25) = 25 \% 10 = 5$

Key = 7 : $H(7) = 7 \% 10 = 7$

Key = 56 : $H(56) = 56 \% 10 = 6$

Key = 95 : $H(95) = 95 \% 10 = 5$

Key = 43 : $H(43) = 43 \% 10 = 3$

Key = 8 : $H(8) = 8 \% 10 = 8$

Key = 13 : $H(13) = 13 \% 10 = 3$

Phương pháp địa chỉ mở

Dò tuyến tính - Linear probing

- Hàm băm: $H(\text{key}) = \text{key} \% M$
- Giải quyết xung đột: $H'(\text{key}) = (H(\text{key}) + i) \% M$

Với $i = 1, 2, \dots, M-1$ (M là kích thước bảng băm)

- Ví dụ: Thêm các key 15, 23, 78, 25, 7, 56, 95, 43, 8, 13 vào bảng băm. Giải quyết xung đột bằng các phương pháp dò tuyến tính
Cho kích thước bảng băm $M = 10$

Dò tuyến tính - Linear probing

Key = {15, 23, 78, 25, 7, 56, 95, 43, 8, 13} M=10

0	NULL	NULL
1	NULL	NULL
2	NULL	NULL
3	23	23
4	NULL	NULL
5	15	15
6	NULL	25
7	NULL	7
8	78	78
9	NULL	56

Key = 15 : $H(15) = 15 \% 10 = 5$

Key = 23 : $H(23) = 23 \% 10 = 3$

Key = 78 : $H(78) = 78 \% 10 = 8$

Key = 25 : $H(25) = 25 \% 10 = 5$ (đụng độ)

Tại $i=1$: $H'(25) = (5+1) \% 10 = 6$

Key = 7 : $H(7) = 7 \% 10 = 7$

Key = 56 : $H(56) = 56 \% 10 = 6$ (đụng độ)

Tại $i=1$: $H'(56) = (6+1) \% 10 = 7$ (đụng độ)

Tại $i=2$: $H'(56) = (6+2) \% 10 = 8$ (đụng độ)

Tại $i=3$: $H'(56) = (6+3) \% 10 = 9$

Dò tuyến tính - Linear probing

Key = {15, 23, 78, 25, 7, 56, 95, 43, 8, 13} M=10

0	NULL	NULL	95
1	NULL	NULL	NULL
2	NULL	NULL	NULL
3	23	23	23
4	NULL	NULL	NULL
5	15	15	15
6	NULL	25	25
7	NULL	7	7
8	78	78	78
9	NULL	56	56

Key = 95 : $H(95) = 95 \% 10 = 5$

Tại $i=1$: $H'(95) = (5+1) \% 10 = 6$

Tại $i=2$: $H'(95) = (5+2) \% 10 = 7$

Dò tuyến tính - Linear probing

Key = {15, 23, 78, 25, 7, 56, 95, 43, 8, 13} M=10

0	NULL	NULL	95	95
1	NULL	NULL	NULL	NULL
2	NULL	NULL	NULL	NULL
3	23	23	23	23
4	NULL	NULL	NULL	43
5	15	15	15	15
6	NULL	25	25	25
7	NULL	7	7	7
8	78	78	78	78
9	NULL	56	56	56

Key = 43 : $H(43) = 43 \% 10 = 3$

Tại $i=1$: $H'(43) = (3+1) \% 10 = 4$

Dò tuyến tính - Linear probing

Key = {15, 23, 78, 25, 7, 56, 95, 43, 8, 13} M=10

0	95
1	8
2	13
3	23
4	43
5	15
6	25
7	7
8	78
9	56

Key = 8 : $H(8) = 8 \% 10 = 8$ (đụng độ)

Tại $i=1$: $H'(8) = (8+1) \% 10 = 9$ (đụng độ)

Tại $i=2$: $H'(8) = (8+2) \% 10 = 0$ (đụng độ)

Tại $i=3$: $H'(8) = (8+3) \% 10 = 1$

Key = 13 : $H(13) = 13 \% 10 = 3$ (đụng độ)

Tại $i=1$: $H'(13) = (3+1) \% 10 = 4$

Tại $i=2$: $H'(13) = (3+2) \% 10 = 5$

Tại $i=3$: $H'(13) = (3+3) \% 10 = 6$

Tại $i=4$: $H'(13) = (3+4) \% 10 = 7$

Tại $i=5$: $H'(13) = (3+5) \% 10 = 8$

Tại $i=6$: $H'(13) = (3+6) \% 10 = 9$

Tại $i=7$: $H'(13) = (3+7) \% 10 = 0$

Tại $i=8$: $H'(13) = (3+8) \% 10 = 1$

Phương pháp địa chỉ mở

Dò bậc hai - Quadratic probing

- Hàm băm: $H(\text{key}) = \text{key} \% M$
- Giải quyết xung đột: $H'(\text{key}) = (H(\text{key}) + i^2) \% M$

Với $i = 1, 2, \dots, M-1$ (M là kích thước bảng băm)

- Ví dụ: Thêm các key 15, 23, 78, 25, 7, 56, 95, 43, 8, 13 vào bảng băm. Giải quyết xung đột bằng các phương pháp dò bậc 2

Cho kích thước bảng băm $M = 10$

Dò bậc hai - Quadratic probing

Key = {15, 23, 78, 25, 7, 56, 95, 43, 8, 17} M=10

0	NULL	NULL
1	NULL	NULL
2	NULL	NULL
3	23	23
4	NULL	NULL
5	15	15
6	NULL	25
7	NULL	7
8	78	78
9	NULL	NULL

Key = 15 : $H(15) = 15 \% 10 = 5$

Key = 23 : $H(23) = 23 \% 10 = 3$

Key = 78 : $H(78) = 78 \% 10 = 8$

Key = 25 : $H(25) = 25 \% 10 = 5$

Tại $i=1$: $H'(25) = (5+1^2) \% 10 = 6$

Key = 7 : $H(7) = 7 \% 10 = 7$

Key = 56 : $H(56) = 56 \% 10 = 6$

Dò bậc hai - Quadratic probing

Key = {15, 23, 78, 25, 7, 56, 95, 43, 8, 17} M=10

0	NULL	NULL	56	56	56
1	NULL	NULL	NULL	NULL	NULL
2	NULL	NULL	NULL	NULL	NULL
3	23	23	23	23	23
4	NULL	NULL	NULL	NULL	43
5	15	15	15	15	15
6	NULL	25	25	25	25
7	NULL	7	7	7	7
8	78	78	78	78	78
9	NULL	NULL	NULL	95	95

Key = 95 : $H(95) = 95 \% 10 = 5$

Tại $i=1$: $H'(95) = (5+1^2) \% 10 = 6$

Tại $i=2$: $H'(95) = (5+2^2) \% 10 = 9$

Key = 43 : $H(43) = 43 \% 10 = 3$

Tại $i=1$: $H'(43) = (3+1^2) \% 10 = 4$

Dò bậc hai - Quadratic probing

Key = {15, 23, 78, 25, 7, 56, 95, 43, 8, 17} M=10

0	56
1	17
2	8
3	23
4	43
5	15
6	25
7	7
8	78
9	95

Key = 8 : $H(8) = 8 \% 10 = 8$

Tại $i=1$: $H'(8) = (8+1^2) \% 10 = 9$

Tại $i=1$: $H'(8) = (8+2^2) \% 10 = 2$

Key = 17 : $H(17) = 17 \% 10 = 7$

Tại $i=1$: $H'(17) = (7+1^2) \% 10 = 8$

Tại $i=2$: $H'(17) = (7+2^2) \% 10 = 1$

Phương pháp địa chỉ mở

Băm kép - Double hashing

- Hàm băm: $H_1(\text{key}) = \text{key} \% M$
- Giải quyết xung đột: $H'(\text{key}) = (H_1(\text{key}) + i * H_2(\text{key})) \% M$

Trong đó: $H_2(\text{key}) = (M-2) - \text{key} \% (M-2)$

Với $i = 1, 2, \dots, M-1$ (M là kích thước bảng băm)

Băm kép - Double hashing

- Ví dụ: Thêm các key 15, 23, 78, 25, 7, 56, 95, 41, 8, 10 vào bảng băm. Giải quyết xung đột bằng các phương pháp dò băm kép
Cho kích thước bảng băm $M = 10$

Băm kép - Double hashing

Key = {15, 23, 78, 25, 7, 56, 95, 41, 8, 10} M=10

0	NULL	NULL	NULL	10
1	NULL	NULL	NULL	41
2	NULL	25	25	25
3	23	23	23	23
4	NULL	NULL	NULL	8
5	15	15	15	15
6	NULL	56	56	56
7	NULL	7	7	7
8	78	78	78	78
9	NULL	NULL	95	95

Key = 15 : $H(15) = 15 \% 10 = 5$

Key = 23 : $H(23) = 23 \% 10 = 3$

Key = 78 : $H(78) = 78 \% 10 = 8$

Key = 25 : $H(25) = 25 \% 10 = 5$

Key = 7 : $H(7) = 7 \% 10 = 7$

Key = 56 : $H(56) = 56 \% 10 = 6$

Key = 95 : $H(95) = 95 \% 10 = 5$

Key = 43 : $H(41) = 41 \% 10 = 1$

Key = 8 : $H(8) = 8 \% 10 = 8$

Key = 10 : $H(10) = 10 \% 10 = 0$

GIẢI QUYẾT ĐỤNG ĐỘ

Phương pháp		Hàm băm	Giải quyết xung đột
Nối kết	Nối kết trực tiếp - Direct Chaining	$H(\text{key}) = \text{key} \% M$	Gom thành 1 dslk
	Nối kết hợp nhất - Coalesced Chaining	$H(\text{key}) = \text{key} \% M$	Nút trống phía cuối mảng
Địa chỉ mở	Dò tuyến tính - Linear probing	$H(\text{key}) = \text{key} \% M$	$H'(\text{key}) = (H(\text{key}) + i) \% M$
	Dò bậc 2 - Quadratic probing	$H(\text{key}) = \text{key} \% M$	$H'(\text{key}) = (H(\text{key}) + i^2) \% M$
	Băm kép - Double hashing	$H_1(\text{key}) = \text{key} \% M$ $H_2(\text{key}) = 7 - \text{key} \% 7$	$H'(\text{key}) = (H_1(\text{key}) + i * H_2(\text{key})) \% M$
$i = 1, 2, \dots, M-1$ (M là kích thước bảng băm)			

Băm lại (rehashing)

Băm lại (rehashing) cần được thực hiện khi:

- Không thể thực hiện một thao tác thêm phần tử mới bất kỳ.
- Bảng băm có hệ số tải đạt một giá trị xác định ($\lambda \geq 0.5$) làm độ phức tạp thời gian của thao tác tìm kiếm tăng.

Băm lại (rehashing)

Sử dụng phương pháp băm key để giải quyết đụng độ

Key = {78, 23, 7, 56, 95, 43} M=10

$H = \text{key} \% 10$

$H'(\text{key}) = (H1(\text{key}) + i * H2(\text{key})) \% M$

Trong đó:

$H1(\text{key}) = \text{key} \% M$

$H2(\text{key}) = (M-2) - \text{key} \% (M-2)$

0	
1	
2	23
3	
4	
5	95
6	56
7	7
8	78
9	

Băm lại (rehashing)

Sử dụng phương pháp băm key để giải quyết đụng độ

Key = {78, 23, 7, 56, 95, 43} M=10

$H(43) = 43 \% 10 = 3$ (xung đột)

$H_1(43) = 43 \% 10 = 3$

$H_2(43) = 8 - 43 \% 8 = 5$

Tại $i=1$:

$H'(43) = (H_1(43) + H_2(43)) \% 10 = 8$ (xung đột)

Tại $i=2$:

$H'(43) = (H_1(43) + 2 * H_2(43)) \% 10 = 3$ (xung đột)

Tại $i=3$:

$H'(43) = (H_1(43) + 3 * H_2(43)) \% 10 = 8$ (xung đột)

Tại $i=4$:

$H'(43) = (H_1(43) + 4 * H_2(43)) \% 10 = 8$ (xung đột)

....

0	
1	
2	23
3	
4	
5	95
6	56
7	7
8	78
9	

Băm lại (rehashing)

Các bước thực hiện khi băm lại:

- Tăng kích thước bảng băm lên m' với m' là một số nguyên tố và $m' \approx 2 * m$.
- Cập nhật hàm băm.
- Băm lại các phần tử có trong bảng băm cũ và thêm vào bảng băm mới.

Băm lại (rehashing)

Sử dụng phương pháp băm key để giải quyết đụng độ

Key = {78, 23, 7, 56, 95, 43} M'=23

$H = \text{key} \% 23$

$H'(\text{key}) = (H1(\text{key}) + i * H2(\text{key})) \% 23$

Trong đó:

$H1(\text{key}) = \text{key} \% 23$

$H2(\text{key}) = 21 - \text{key} \% 21$

0	
1	
2	
3	
4	
...	
...	
21	
22	
23	