

ĐHSP

Bảng băm (Hash table)



Nội dung

- 1 Đặt vấn đề
- 2 Bảng băm
- 3 Hàm băm
- 4 Các phương pháp xử lý đụng độ



Đặt vấn đề

VCD là một công ty điện thoại lớn, và họ muốn cung cấp cho khách hàng dịch vụ sau:

- Từ một số điện thoại, cho biết tên của thuê bao.
- Các số điện thoại nằm trong khoảng từ $0 - 10^7 - 1$



Cần phải tổ chức lưu trữ như thế nào để việc tìm kiếm một số điện thoại cho trước ít tốn công sức nhất trong giới hạn bộ nhớ cho phép

20/06/2010



Đặt vấn đề - Cách giải quyết 1

❖ Dùng mảng với các index chính là các số điện thoại. Chi phí tìm kiếm là $O(1)$, nhưng chi phí lưu trữ quá lớn – có quá nhiều vùng nhớ lãng phí.

(null)	(null)	...	Lê Lợi	...	(null)
000 - 000	000 - 0001	...	865 - 3999	...	999 - 9999

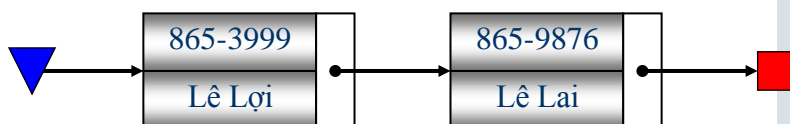
20/06/2010





Đặt vấn đề - cách giải quyết 2

- ❖ Dùng danh sách liên kết, chi phí tìm kiếm là $O(n)$, bộ nhớ chiếm $O(n)$



20/06/2010



Đặt vấn đề - cách giải quyết 3

- ❖ Dùng cây nhị phân cân bằng. Chi phí tìm kiếm là $O(\log n)$, bộ nhớ chiếm $O(n)$.
- ❖ Có thể vượt qua ngưỡng $O(\log n)$ với chi phí về bộ nhớ chấp nhận được không?



Dùng bảng băm (hashing table): chi phí tìm kiếm $O(1)$, chi phí lưu trữ $O(N + M)$, M là kích thước của bảng băm.

20/06/2010





Bảng băm

- ❖ Bảng băm tương tự như mảng nhưng kèm theo một hàm băm để ánh xạ nhiều giá trị vào cùng một phần tử trong mảng.
- ❖ Chẳng hạn, dùng hàm modulo: vị trí chứa khóa k sẽ là $k \bmod M$ (M là kích thước mảng băm).

20/06/2010



Hàm băm

- ❖ Là hàm biến đổi khóa k của phần tử thành địa chỉ trong bảng băm.
- ❖ Là một ánh xạ thích hợp từ tập các khóa U vào tập các địa chỉ A

$$h: U \longrightarrow A$$

$$K \longrightarrow a = h(k)$$

- Cần tính toán nhanh
- Các khóa phân bố đều
- Ít xảy ra đụng độ

20/06/2010





Ví dụ

❖ Cho các khóa: 10, 35, 302, 23, 15, 74, 57, 19, 41.

Với hàm băm $h(k) = k \bmod 10$.

20/06/2010



Hàm băm

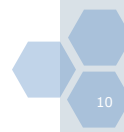
❖ Khi cần băm các khóa không nguyên thì phải làm thế nào?

- Tìm cách chuyển thành các giá trị nguyên

Vd: Trong trường hợp chúng ta, loại bỏ dấu gạch nối: 865-3999 thành 8653999. Nếu các khóa là chuỗi ký tự: sử dụng mã ASCII.

- Sau đó dùng hàm băm nguyên bình thường.

20/06/2010





Bảng băm – ví dụ (tt)

❖ Thêm Lê Lai (865-9876)

❖ Chuyện gì xảy ra nếu thêm Anh Đức (863-2234)?

(null)	Lê Lai	(null)	(null)	Lê Lợi
0	1	2	3	4

20/06/2010

11



Giải quyết đụng độ

❖ Phương pháp nối kết (Chaining)

❖ Phương pháp địa chỉ mở (Open-addressing)

- Phương pháp dò tuyến tính (Linear probing)
- Phương pháp dò bậc 2 (Quadratic probing)
- Phương pháp băm kép (Double hashing)

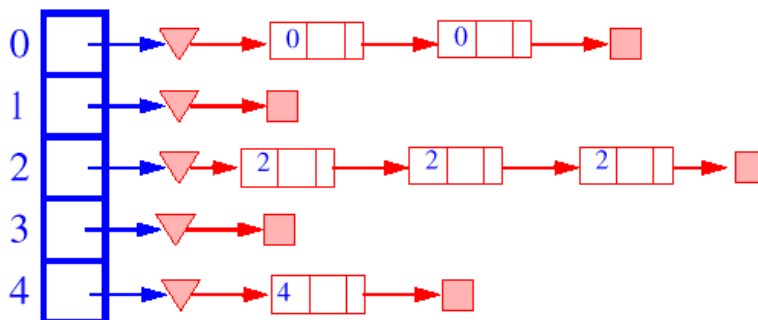
20/06/2010

12



Phương pháp nối kết

Xâu các phần tử có “**khóa**” được băm vào **cùng một địa chỉ** thành một danh sách liên kết → Bảng băm là một mảng các danh sách liên kết.



20/06/2010

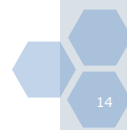


PP nối kết - ví dụ

❖ Cho các khóa: 10, 35, 30, 21, 39, 23, 16, 74, 57, 19, 41, 42.

Với hàm băm $h(k) = k \bmod 7$.

20/06/2010





PP nối kết - Ví dụ

Ký tự thứ i trong bảng chữ cái được biểu diễn bởi số i , dùng hàm băm $h(k) = k \bmod 11$ và phương pháp kết nối để xử lý đụng độ thì kết quả như sau:

Key: A S E A R C H I N G E X A M P L E
Hash: 1 8 5 1 7 3 8 9 3 7 5 2 1 2 5 1 5

0	1	2	3	4	5	6	7	8	9	10
	A	M	C		E		G	H	I	
	A	X	N		E		R	S		
	A				E					
	L				P					



Phương pháp địa chỉ mở

❖ Ý tưởng:

Nếu vị trí hiện tại đã bị khóa khác chiếm, ta sẽ thử tìm đến vị trí kế tiếp nào đó trong bảng.

❖ Dò tuyến tính:

$$H(k, i) = (H(k, i) + i) \bmod M$$

❖ Băm kép:

$$H(k, i) = (H(k) + i * H_2(k)) \bmod M$$

i : thứ tự của lần thử ($i = 0, 1, 2, \dots$)

$H(k)$ và $H_2(k)$: hàm băm

M : số phần tử của bảng băm

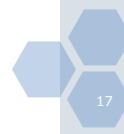


Dò tuyến tính – thuật toán

```

linear_probing_insert(K)
  if (table is full) error
  probe = h(K)
  while (table[probe] occupied)
    probe = (probe + 1) mod M
  table[probe] = K
  
```

20/06/2010



Dò tuyến tính – ví dụ

Cho các khóa: 10, 33, 302, 23, 15, 74, 53

Với $h(k) = k \bmod 10$

0	1	2	3	4	5	6	7	8	9
10									
10			33						
10		302	33						
10		302	33	23					
10		302	33	23	15				
10		302	33	23	15	74			
10		302	33	23	15	74	53		

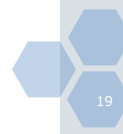
20/06/2010



Dò tuyến tính – Đánh giá

- ❖ Ít tốn bộ nhớ hơn dùng danh sách liên kết (chaining)
- ❖ Chậm hơn dùng danh sách liên kết (có thể phải duyệt dọc theo bảng trên con đường dài)
- ❖ Việc hủy một phần tử khó khăn:
 - Hoặc đánh dấu ô tương ứng là “xóa”
 - Hoặc chuyển một số phần tử ngược lên trước để điền vào ô bị xóa.

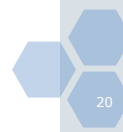
20/06/2010



Băm kép – Thuật toán

```
double_hash_insert(K)
  if(table is full) error
  probe = h1(K)
  offset = h2(K)
  while (table[probe] occupied)
    probe = (probe + offset) mod M
  table[probe] = K
```

20/06/2010





Băm kép – Ví dụ

- ❖ $h = k \bmod 13$
 $d = 7 - (k \bmod 7)$
- ❖ Lần lượt chèn các khóa sau vào bảng:
18, 41, 22, 44, 59, 32,
31, 73

k	$h(k)$	$d(k)$	Probes	
18	5	3	5	
41	2	1	2	
22	9	6	9	
44	5	5	5	10
59	7	4	7	
32	6	3	6	
31	5	4	5	9 0
73	8	4	8	

0	1	2	3	4	5	6	7	8	9	10	11	12

↓

31		41				18	32	59	73	22	44	
0	1	2	3	4	5	6	7	8	9	10	11	12

20/06/2010



Băm kép – Đánh giá

- ❖ Có ưu khuyết điểm giống thăm dò tuyến tính.
- ❖ Phân bố các khóa đều hơn so với thăm dò tuyến tính.
- ❖ Cần chọn hàm băm thích hợp để không bị lặp vô tận.

20/06/2010





Các chức năng cơ bản

- ❖ **Init:** Khởi tạo một bảng băm với m danh sách rỗng
- ❖ **Insert:** $\text{index} = h(k)$
Chèn vào bảng tại vị trí $\text{Table}[\text{index}]$
- ❖ **Find:** $\text{index} = h(k)$
Lần theo danh sách tại $\text{Table}[\text{index}]$, tìm k và trả về thông tin tìm thấy (hoặc báo lỗi nếu ko tìm thấy).
- ❖ **Remove:** $\text{index} = h(k)$
Lần theo danh sách tại $\text{Table}[\text{index}]$, tìm k và huỷ thông tin tìm thấy (hoặc báo lỗi nếu ko tìm thấy).

20/06/2010

23



Băm kép – Bài tập

- ❖ Xét các khóa sau:
13, 33, 29, 15, 50, 74, 20, 101
- ❖ Cho biết kết quả của bảng băm khi dùng các kỹ thuật sau để giải quyết đụng độ: (dùng hàm băm $h = k \bmod 7$)
 - Dùng pp nối kết
 - Dùng pp thăm dò tuyến tính
 - Dùng hàm băm kép: $h1 = k \bmod 7$
 $h2 = 5 - (k \bmod 5)$

20/06/2010

24

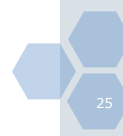


Bài tập

❖ Cho các khóa: E A S Y Q U E S T I O N

- Dùng hàm băm $h(k) = k \bmod 13$ với kỹ thuật dò tuyến tính
- Dùng hàm băm kép
 $h_1(k) = k \bmod 13; h_2 = 1 + (k \bmod 11)$

20/06/2010



ĐHSP



Thank You!

