

Tìm kiếm theo bảng băm

Hash Table

Cấu trúc dữ liệu và giải thuật – HCMUS 2011

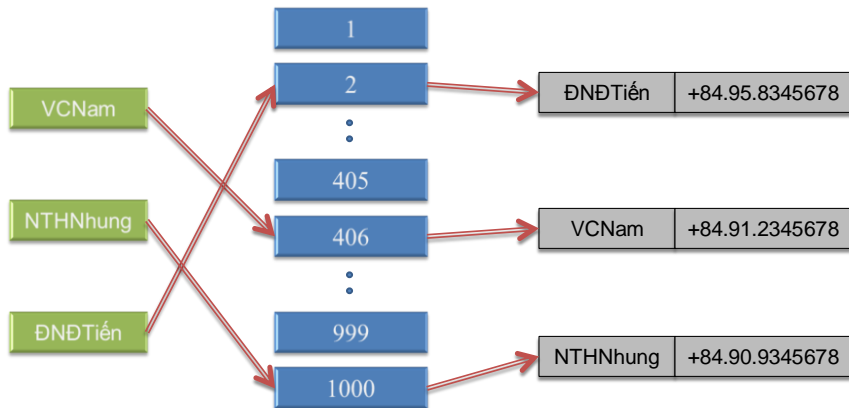
Khái quát về hash

- ◉ **Vấn đề:** Cho trước 1 tập S gồm các phần tử được đặc trưng bởi giá trị khóa. Trên giá trị các khóa này có quan hệ thứ tự. Tổ chức S như thế nào để tìm kiếm 1 phần tử có khóa k cho trước có độ phức tạp ít nhất trong giới hạn bộ nhớ cho phép?
- ◉ **Ý tưởng:** Biến đổi khóa k thành một số (bằng hàm hash) và sử dụng số này như là địa chỉ để tìm kiếm trên bảng dữ liệu.

Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Ví dụ về một bảng băm

33



Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Độ phức tạp

34

- Chi phí tìm kiếm trung bình: $O(1)$
- Chi phí tìm kiếm trong trường hợp xấu nhất: $O(n)$ (rất ít gặp).

Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Hàm băm (hash function)

35

- ◉ **Định nghĩa:** là hàm biến đổi khóa k của phần tử thành địa chỉ trong *bảng băm*.

Ví dụ: $H(k) = k \bmod M$.

- ◉ **Tổng quát về phép biến đổi khóa:** Là 1 ánh xạ thích hợp từ tập các khóa U vào tập các địa chỉ A .

$$H: U \rightarrow A$$

$$k \rightarrow a = h(k)$$

Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Khó khăn của hàm băm

36

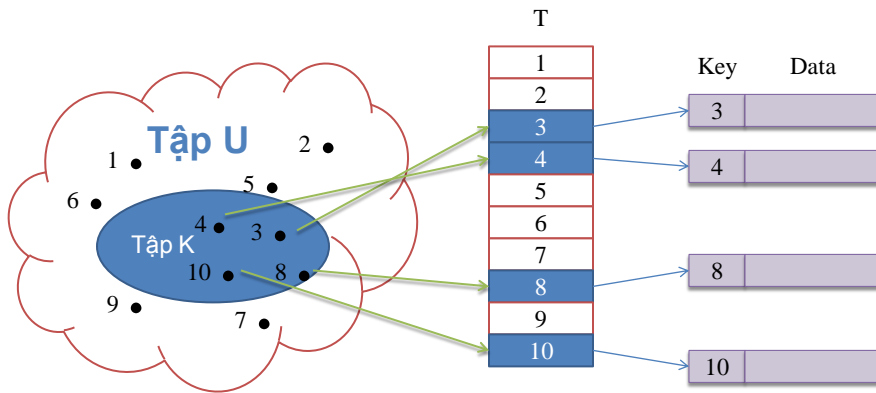
- ◉ Tập các giá trị khóa (U) có thể lớn hơn rất nhiều so với số khóa thực tế (K) rất nhiều.
- ◉ Ví dụ: Quản lý danh sách 1000 sinh viên, mã sinh viên gồm 7 chữ số.

Có $U = 10^7$ khóa so với $K = 1000$.

Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Ví dụ

37



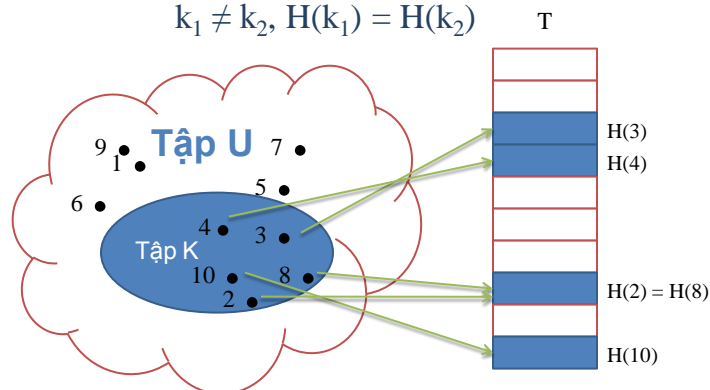
Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Sự đụng độ (collision)

38

◉ $\exists k_1, k_2 \in K:$

$$k_1 \neq k_2, H(k_1) = H(k_2)$$



Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Những yêu cầu đối với hàm băm

39



Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Ví dụ về hàm băm

40

- ◉ Xét lại ví dụ về danh sách sinh viên:
Với kích thước bảng là $M = 1000$, ta có thể chọn hàm băm như sau:
$$H(k) = k \bmod M.$$
- ◉ Khóa này thỏa mãn yêu cầu tính toán nhanh và trải đều trên bảng.

Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Các phương pháp xử lý đụng độ

41

- Phương pháp nối kết (chaining)
- Phương pháp địa chỉ mở (Open-addressing)

Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Phương pháp nối kết

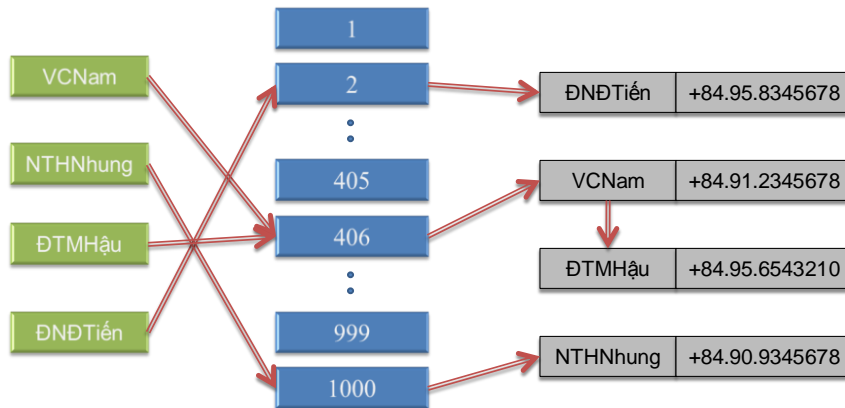
42

- Ứng với mỗi địa chỉ của bảng, ta có một danh sách liên kết chứa các phần tử có khóa khác nhau mà có cùng địa chỉ đó.
- Ta sẽ có danh sách (bảng băm) gồm M phần tử chứa địa chỉ đầu của các danh sách liên kết.

Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Giải quyết đụng độ với phương pháp nối kết

43



Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Phương pháp địa chỉ mở

44

- Tên gọi khác:
 - ▣ Phương pháp dò
 - ▣ Phương pháp thử
- Ý tưởng:
 - ▣ Khi đụng độ xảy ra, ta sẽ thử tìm đến vị trí kế tiếp nào đó trong bảng cho đến khi tìm thấy vị trí nào còn trống.

Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Các cách thực hiện

45

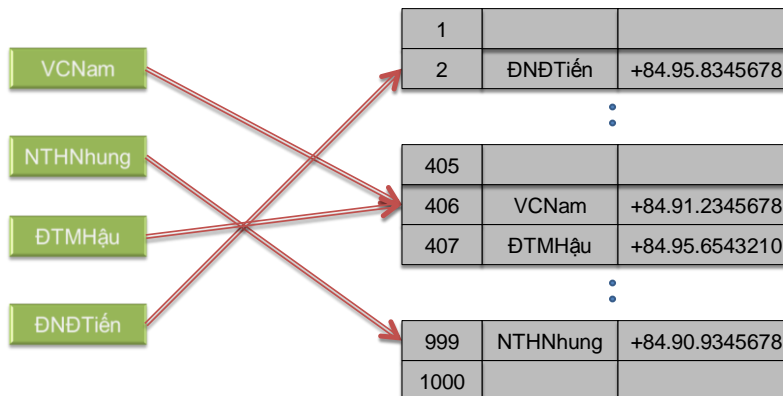
- Phương pháp dò tuyến tính (Linear probing)
- Phương pháp dò bậc 2 (Quadratic probing)
- Phương pháp băm kép (Double hashing)

Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Giải quyết đụng độ bằng phương pháp dò tuyến tính

46

- Ý tưởng: $H(k, i) = (h(k) + i) \bmod M$



Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Phương pháp dò bậc 2 và băm kép

47

- ◉ *Phương pháp dò bậc 2:*

$$H(k, i) = (h(k) + i^2) \bmod M$$

- ◉ *Phương pháp băm kép:*

$$H(k, i) = (h_1(k) + i \cdot h_2(k)) \bmod M$$

Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Ưu thế của phương pháp địa chỉ mở so với phương pháp nối kết

48

- ◉ Đơn giản khi cài đặt.
- ◉ Sử dụng cấu trúc dữ liệu cơ bản.
- ◉ Phương pháp địa chỉ mở giải quyết được đụng độ nhưng lại có thể gây ra đụng độ mới.
- ◉ Phương pháp nối kết không bị ảnh hưởng về tốc độ khi mảng gần đầy.
- ◉ Ít tốn bộ nhớ khi mảng thừa (ít phần tử).

Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Bài tập

49

1. Cho bảng băm có kích thước $M = 11$. Hàm băm: $h(k) = k \bmod M$. Dùng phương pháp địa chỉ mở. Cho biết kết quả sau khi thêm vào bảng băm các khóa 10, 22, 31, 4, 15, 28, 17, 88, 59, với 3 phương pháp xử lý đụng độ:
 - a. Dò tuyến tính.
 - b. Dò bậc 2.
 - c. Băm kép $h_2(k) = (k \bmod 19) + 1$.

Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Bài tập

50

2. Cho từ điển Anh – Việt có 15.000 từ, hãy tổ chức cấu trúc dữ liệu bảng băm và cho biết hàm băm thích hợp giúp cho việc tra từ hiệu quả nhất.

Cấu trúc dữ liệu và giải thuật – HCMUS 2011

Hỏi và Đáp

Cấu trúc dữ liệu và giải thuật – HCMUS 2011