

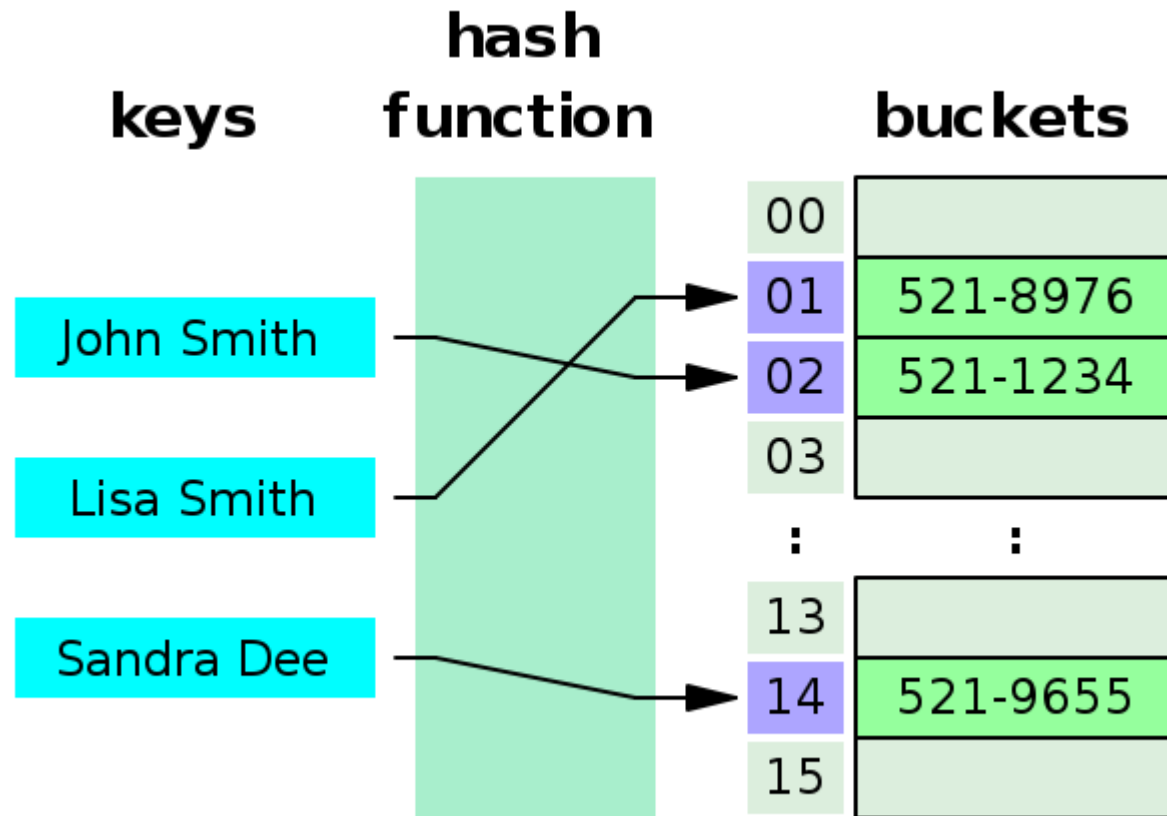


Bảng băm

L/O/G/O

www.themegallery.com

➡ Lý thuyết về bảng băm



Hàm băm

- Hàm băm là bất kỳ hàm nào có thể được sử dụng để ánh xạ tập dữ liệu có kích thước tùy ý thành tập dữ liệu có kích thước cố định và đưa vào bảng băm. Các giá trị được trả về bởi hàm băm được gọi là giá trị băm.

Hàm băm

Một hàm băm được đánh giá tốt nếu nó đạt được các yêu cầu cơ bản sau:

- Dễ tính toán: Nó phải dễ tính toán và bản thân nó không phải là một thuật toán
- Phân bố đồng đều: Nó cần phải phân phối đồng đều trên bảng băm, không xảy ra việc tập trung thành các cụm
- Ít va chạm: Va chạm xảy ra khi các cặp phần tử được ánh xạ tới cùng một giá trị băm



Hàm Băm sử dụng Phương pháp chia

- Dùng số dư:
 - $h(k) = k \bmod m$
 - k là khoá, m là kích thước của bảng.
- Vấn đề chọn giá trị m
 - $m = 2^n$ (không tốt)
 - nếu chọn $m = 2^n$ thông thường không tốt $h(k) = k \bmod 2^n$ sẽ chọn cùng n bits cuối của k
 - m là nguyên tố (tốt). Thông thường m được chọn là số nguyên tố gần với 2^n . Chẳng hạn bảng ~4000 mục, chọn $m = 4093$

➡ Hàm Băm sử dụng Phương pháp nhân

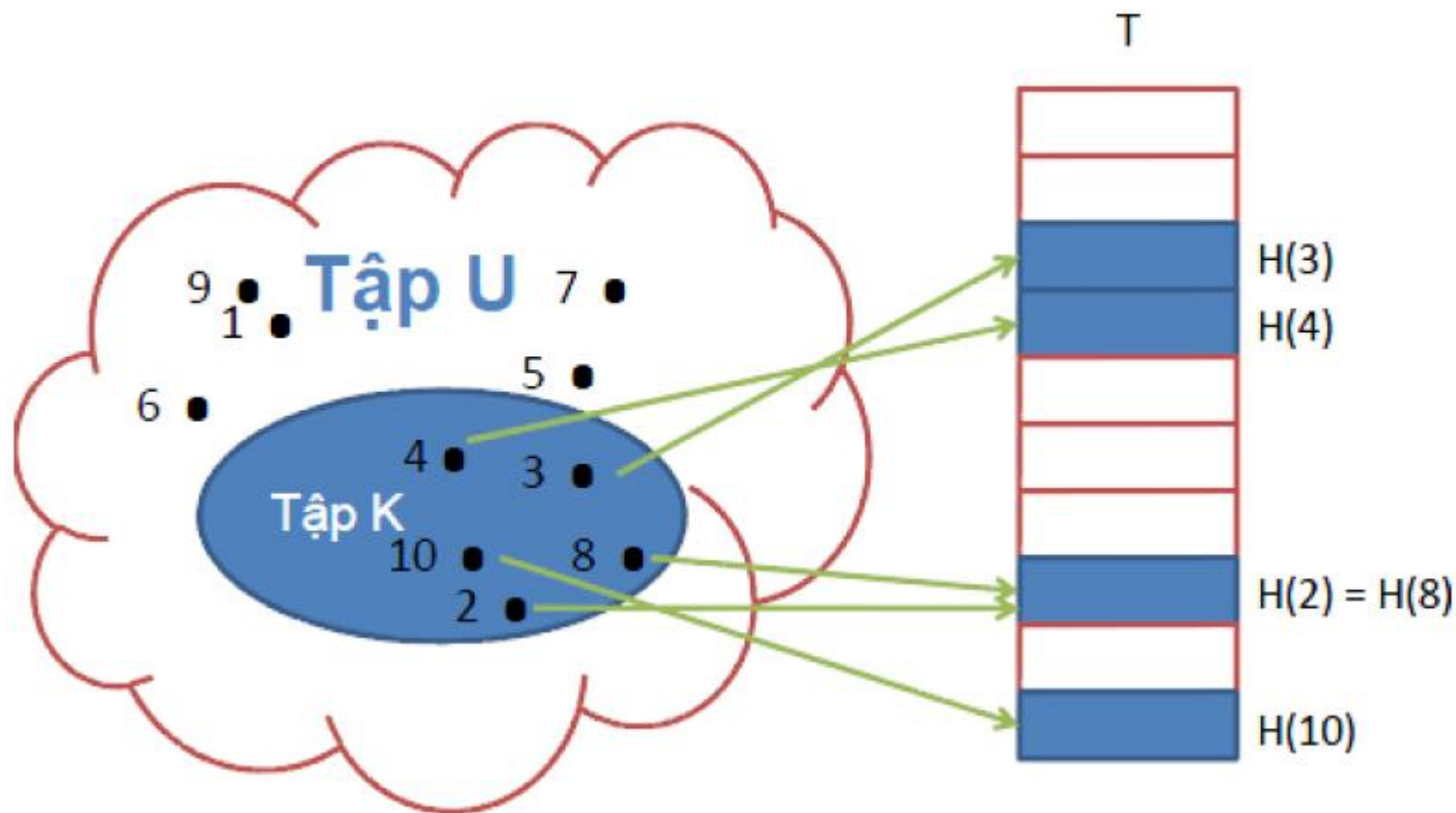
- Sử dụng
 - $h(k) = m * (k * A \bmod 1)$
 - k là khóa, m là kích thước bảng, A là hằng số: $0 < A < 1$
- Chọn m và A
 - thường chọn $m = 2^p$
 - Sự tối ưu trong việc chọn A phụ thuộc vào đặc trưng của dữ liệu.
 - Theo Knuth chọn $A = 1/2(\sqrt{5} - 1) \approx 0.618033987$ được xem là tốt.

➡ Phép băm phổ quát

- Việc chọn hàm băm m không tốt có thể dẫn đến xác suất đụng độ lớn.
- Giải pháp:
 - Lựa chọn hàm băm h ngẫu nhiên.
 - Chọn hàm băm độc lập với khóa.
 - Khởi tạo một tập các hàm băm H phổ quát và từ đó h được chọn ngẫu nhiên.
 - Một tập các hàm băm H là phổ quát (*universal*) nếu với mọi f , k belong H và 2 khóa k, l ta có xác suất: $\Pr\{f(k) = f(l)\} \leq 1/m$

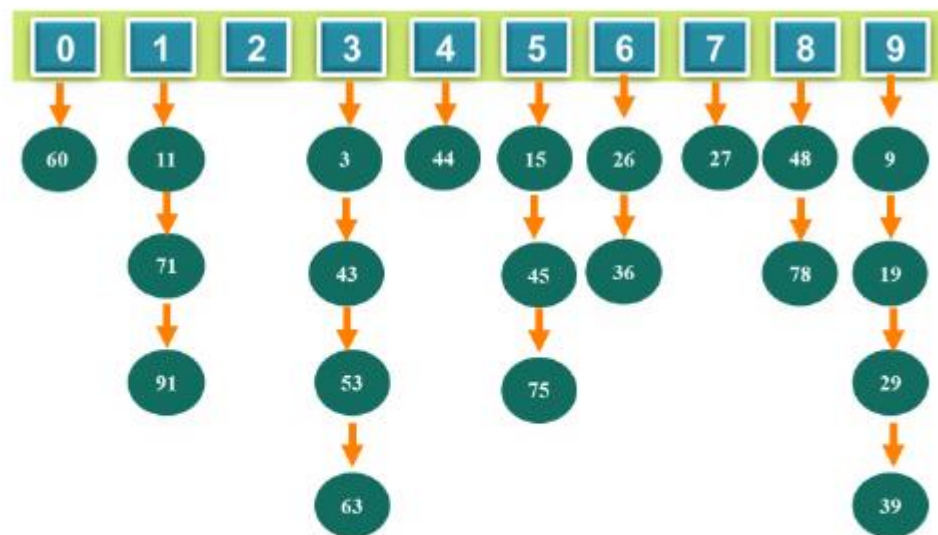
➔ Xung đột

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



➡ Giải quyết độ

- Phương pháp kết nối trực tiếp



K: tập các số tự nhiên

$M = \{0, 1, 2, \dots, 9\}$

$F(h) = k \text{ MOD } M$

➡ Giải quyết độ

Phương pháp đo tuyến tính

- Hàm băm lại lần i được biểu diễn bằng công thức sau:
- $f(key) = (f(key) + i) \% M$
- Trong đó $f(key)$ là hàm băm chính của bảng băm.

Giải quyết đụng độ

Tương tự dò tuần tự nhưng hàm băm lại lần thứ i được biểu diễn bằng công thức sau:

- $fi(key) = (f(key) + i^2) \% M$
- Trong đó: $f(key)$ là hàm băm chính của bảng băm.
- Nếu đã dò đến cuối bảng thì trở về dò lại từ đầu bảng.

Giải quyết đụng độ

- Phương pháp băm kép:
- $h1(key) = key \% M$.
- $h2(key) = (M-2) - key \% (M-2)$.

$$H(key, i) = (h1(k) + i * h2(k)) \text{ Mod } M$$

Bài tập

Bài 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 $h_2(k) = (k \bmod 19) + 1$.



Bài 2: Viết lại chương trình(GV cung cấp), cho phép người dùng tự nhập số lượng giá trị k và thông tin của SV cần đưa vào bảng băm.



Bài 3: Áp dụng các phương pháp sau để giải quyết độ

- Phương pháp nối kết
- Phương pháp dò tuyến tính
- Phương pháp dò bậc 2
- Phương pháp bấm kép