Nguyễn Minh Đức 1712358

Phần trắc nghiệm

1. B
2. B
3. D
4. D
5. D
6. A
7. A
8. D
9. B
10. C

Phần tự luận

1.

Dùng hàm băm h để băm n khóa (key) vào bảng băm T có m phần tử, => số lần đụng độ (P) có thể dao động trong đoạn [0;n-1].

Coi các key đưa vào là ngẫu nhiên, ta có thể tính được số lần đụng độ trung bình (kỳ vọng).

Gọi X là biến ngẫu nhiên chỉ số phần tử chứa ít nhất 1 key sau khi băm của bảng T.

Gọi X_i là biến ngẫu nhiên sao cho X_i=1 nếu T[i] chứa ít nhất 1 key,

hoặc X_i=0 nếu T[i] không chứa key nào.

=> X = $\sum_{i=0}^{m-1} X_i$

Suy ra kỳ vọng của X sẽ bằng kỳ vọng của tổng trên

E(X) = E($\sum_{i=0}^{m-1} X_i$) = $\sum_{i=0}^{m-1} E(X_i)$

Tính E(X_i)

$$E(X_i) = 0*(\frac{m-1}{m})^n+1*(1 - (\frac{m-1}{m})^n) = 1 - (\frac{m-1}{m})^n$$

Từ đây ta thấy các E(X_i) đều bằng nhau và không phụ thuộc i, có m giá trị X_i

$$=> E(X) = m*(1 - (\frac{m-1}{m})^n) = m – m*(\frac{m-1}{m})^n = m – m*(1 - \frac{1}{m})^n$$

=> kỳ vọng số lần đụng độ P sẽ là

$$E(P) = n – E(X) = n – m + m*(1 - \frac{1}{m})^n$$

Vì ý nghĩa của những con số không được rõ nên ta biến đổi 1 chút

$(1 - \frac{1}{m})^m$ sẽ sấp sỉ $\frac{1}{e}$ nếu m đủ lớn, ở đây nếu m = 10 thì sấp sỉ này chỉ lệch có 5% so với $\frac{1}{e}$, nên với những con số thực tế ta có thể coi $(1 - \frac{1}{m})^m = \frac{1}{e}$

Như vậy:

$$E(P) = n - m + m*(1 - \frac{1}{m})^{\frac{m*n}{m}} = n - m + m*(\frac{1}{e})^{\frac{n}{m}}$$

$$= n - m(1 - \left(\frac{1}{e}\right)^{\frac{n}{m}})$$

Trên đây là kỳ vọng số lần đụng độ khi băm n key vào bảng T m phần tử.

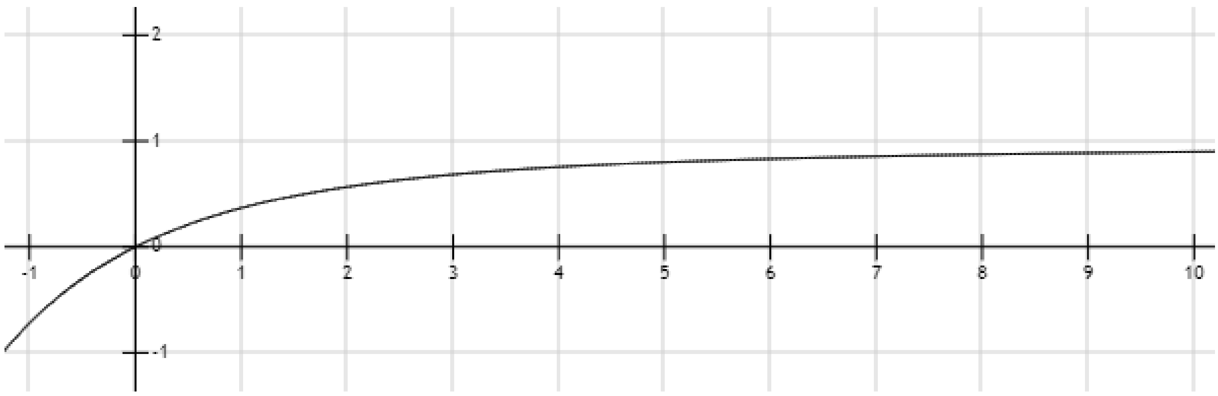
Ta có thể tính vài trường hợp đặc biệt

Khi n = m, khi đó số lần đụng độ được kỳ vọng là $\frac{n}{e}$ gần bằng $\frac{n}{3}$

Nếu đặt hệ số tải k =$\frac{n}{m}$ thì:

$$E(P) = n*(\frac{1}{k}\left(\frac{1}{e}\right)^k - \frac{1}{k} + 1)$$

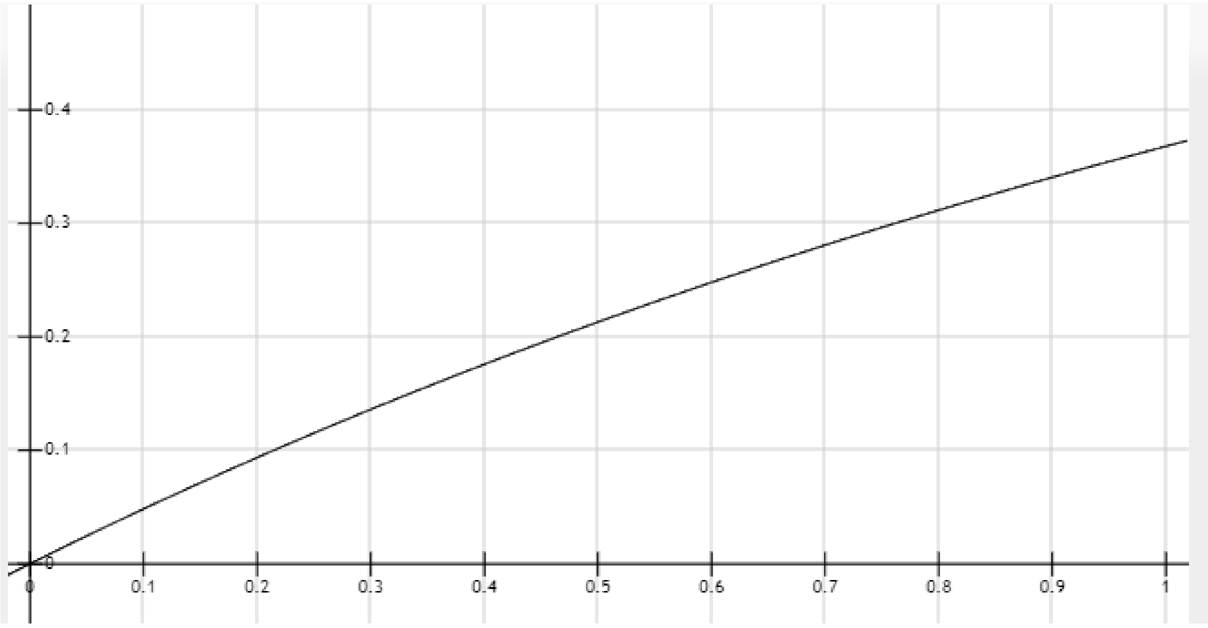
Hàm f(k) = $\frac{1}{k}\left(\frac{1}{e}\right)^k - \frac{1}{k} + 1$ có đồ thị như sau:



Hiển nhiên khi k lớn hàm số trên sẽ tiệm cận 1, số lần đụng độ sẽ tiệm cận n

Ở đây chỉ cần k=3 thì f(k) bằng khoảng 2/3, tức gần 2/3 số key sẽ bị đụng độ, hiệu quả của bẳng băm bị giảm mạnh nếu k>1

Với k < 1 thì đồ thị f(k) như sau



Từ đây ta thấy tỉ lệ tăng số lần đụng sẽ tăng gần như tuyến tính theo k với hệ số góc khoảng 0.5~0.4 (tất nhiên tỉ lệ này sẽ được nhân với n).

Ở đây khi muốn số lần đụng độ chỉ bằng 1/10 số key, thì k sẽ bằng khoảng 0.2, tức bảng băm T sẽ có số phần tử m gấp 5 lần n.

Ví dụ khi băm 100 key vào bảng T 500 phần tử thì số lần đụng độ kỳ vọng vào khoảng 9.365 gần bằng 1/10 số key.

Còn nếu không chấp nhận được số phần tử m lớn quá 2 lần n thì sẽ phải chấp nhận số lần đụng độ vào khoảng n/5.

Ta có nhận xét thêm là sẽ rất khó để có bảng băm mà không xảy ra đụng độ, khi m n đủ lớn thì k phải cực nhỏ mới mới đảm bảo số lần đụng độ nhỏ hơn 1 (hoặc số lần đụng độ kỳ vọng chỉ bằng 1)

Ví dụ khi n = 100, để f(k) < 1/100 thì k sẽ xấp xỉ 0.02, tức bảng băm T sẽ có phần tử lớn hơn 50*100 = 5000 phần tử, và nếu n = 1000 thì khi này f(k) phải nhỏ hơn 1/1000, tức k sẽ xấp xỉ 0.002, tức bảng băm T phải có số phần tử lớn hơn 500*1000= 500000 phần tử.

Vậy sẽ như thế nào nếu n = 1000000, f(k) sẽ gần như là đường thẳng với hệ số góc 0.5 nếu k đủ nhỏ, từ đây ta tính được bảng băm T sẽ phải có số phần tử lớn hơn 500000*1000000 = $500*10^9$, 500 tỉ phần tử, nếu mỗi phần tử là 1 số nguyên kiểu int 4 byte, thì bộ nhớ cần có sẽ lên tới 2 terabyte. (Khi tính ra đến đây em tự nhiên thấy vô cùng khủng khiếp, thấy bảng băm dường như mất đi sự hâm mộ trong lòng).

2.

Thêm 9 phần tử: 5; 28; 19; 15; 20; 33; 12; 17; 10 vào bảng băm

Thêm 5. h(5) = 5

Vị trí	0	1	2	3	4	5	6	7	8
						5			

Thêm 28. h(28) = 1

Vị trí	0	1	2	3	4	5	6	7	8
		28				5			

Thêm 19. h(19) = 1

Vị trí	0	1	2	3	4	5	6	7	8
		19→28				5			

Thêm 15. h(15) = 6

Vị trí	0	1	2	3	4	5	6	7	8
		19→28				5	15		

Thêm 20. h(20) = 2

Vị trí	0	1	2	3	4	5	6	7	8
		19→28	20			5	15		

Thêm 33. h(33) = 6

Vị trí	0	1	2	3	4	5	6	7	8
		19→28	20			5	33→15		

Thêm 12. h(12) = 3

Vị trí	0	1	2	3	4	5	6	7	8
		19→28	20	12		5	33→15		

Thêm 17. h(17) = 8

Vị trí	0	1	2	3	4	5	6	7	8
		19→28	20	12		5	33→15		17

Thêm 10. h(10) = 1

Vị trí	0	1	2	3	4	5	6	7	8
		10→19→28	20	12		5	33→15		17

3.

Bình thường nếu chuỗi đụng độ không được sắp xếp tăng dần thì với độ dài chuỗi là m, thời gian để tìm được phần tử trong chuỗi là O(m), nếu phần tử không có trong chuỗi thì thời gian tìm sẽ là tối đa (m lần so sánh và dịch chuyển). Thời gian thêm 1 phần tử vào chuỗi đụng độ là O(1). Thời gian xóa 1 phần tử O(m) (tìm và xóa)

Nếu chuỗi đụng độ được sắp xếp tăng dần thì thời gian tìm thấy 1 phần tử trong chuỗi là O(m), thời gian thêm phần tử vào chuỗi là O(m) (tìm và thêm), nhưng trong trường hợp phần tử không có trong chuỗi thì có khả năng cao không cần phải di chuyển so sánh hết cả chuỗi, vì nếu gặp phần tử lớn hơn phần tử cần tìm thì quá trình này có thể dừng lại ngay (bởi nếu đi tiếp thì các phần tử phía sau cũng vẫn sẽ lớn hơn phần tử cần tìm). Nhưng nếu phần tử cần tìm lớn hơn tất cả phần tử trong chuỗi thì vẫn phải di chuyển so sánh tất cả các phần tử trong chuỗi. Thời gian xóa 1 phần tử O(m) (tìm và xóa)

4.

Thêm các phần tử sau vào bảng băm: 10, 22, 31, 4, 15, 28, 17, 88, 59

Với hàm băm h(x) = x mod 11

a) Xử lý đụng độ xâu chuỗi

Thêm 10. h(10) = 10

Vị trí	0	1	2	3	4	5	6	7	8	9	10
											10

Thêm 22. h(22) = 0

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22										10

Thêm 31. h(31) = 9

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22									31	10

Thêm 4. h(4) = 4

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22				4					31	10

Thêm 15. h(15) = 4

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22				15→4					31	10

Thêm 28. h(28) = 6

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22				15→4		28			31	1

Thêm 17. h(17) = 6

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22				15→4		17→28			31	10

Thêm 88. h(88) = 0

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	88→22				15→4		17→28			31	10

Thêm 59. h(59) = 4

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	88→22				59→15→4		17→28			31	10

b) Xử lý đụng độ bằng phương pháp dò tuyến tính

Thêm 10. h(10) = 10

Vị trí	0	1	2	3	4	5	6	7	8	9	10
											10

Thêm 22. h(22) = 0

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22										10

Thêm 31. h(31) = 9

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22									31	10

Thêm 4. h(4) = 4

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22				4					31	10

Thêm 15. h(15) = 4

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22				4	15				31	10

Thêm 28. h(28) = 6

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22				4	15	28			31	1

Thêm 17. h(17) = 6

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22				4	15	28	17		31	10

Thêm 88. h(88) = 0

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22	88			4	15	28	17		31	10

Thêm 59. h(59) = 4

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22	88			4	15	28	17	59	31	10

c) Xử lý đụng độ bằng phương pháp dò bậc 2

Thêm 10. h(10) = 10

Vị trí	0	1	2	3	4	5	6	7	8	9	10
											10

Thêm 22. h(22) = 0

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22										10

Thêm 31. h(31) = 9

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22									31	10

Thêm 4. h(4) = 4

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22				4					31	10

Thêm 15. h(15) = 4, đụng độ, bước nhảy $k^2 = 1$

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22				4	15				31	10

Thêm 28. h(28) = 6

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22				4	15	28			31	1

Thêm 17. h(17) = 6, đụng độ, bước nhảy $k^2 = 1$

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22				4	15	28	17		31	10

Thêm 88. h(88) = 0, đụng độ, bước nhảy $k^2 = 1$

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22	88			4	15	28	17		31	10

Thêm 59. h(59) = 4, đụng độ, bước nhảy $k^2 = 4$

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22	88			4	15	28	17	59	31	10

d) Xử lý đụng độ bằng hàm băm kép h’(k) = 1 + k mod (m-1) = 1+(k mod 10)

Thêm 10. h(10) = 10

Vị trí	0	1	2	3	4	5	6	7	8	9	10
											10

Thêm 22. h(22) = 0

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22										10

Thêm 31. h(31) = 9

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22									31	10

Thêm 4. h(4) = 4

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22				4					31	10

Thêm 15. h(15) = 4, đụng độ, h(1,15) = (h(15) + 2*h’(15)) mod 11 = 5

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22				4	15				31	10

Thêm 28. h(28) = 6

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22				4	15	28			31	10

Thêm 17. h(17) = 6, đụng độ, h(1,17) = (h(17) + 1*h'(17)) mod 11 = 3

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22			17	4	15	28			31	10

Thêm 88. h(88) = 0, đụng độ, h(2,88) = (h(88) + 2*h'(88)) mod 11 = 7

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22			17	4	15	28	88		31	10

Thêm 59. h(59) = 4, đụng độ, h(2,59) = (h(59) + 2*h'(59)) mod 11 = 2

Vị trí	0	1	2	3	4	5	6	7	8	9	10
	22		59	17	4	15	28	88		31	10