

Problem ("Dictionary")

Given set S of n integers in $\{0, 1, \dots, U-1\}$
 build a data structure s.t. we can quickly answer
membership queries (given y , is $y \in S$?)

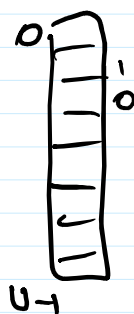
e.g. sorted array $\Rightarrow$ $O(n \log n)$ preproc. time
 $O(n)$ space
 $O(\log n)$ query time

balanced search tree $\Rightarrow$ $O(\log n)$ insert/delete time

better?

Method 0

$O(1)$ time
 $O(U)$ space
 preproc time

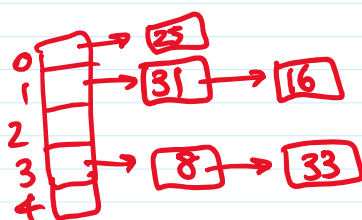
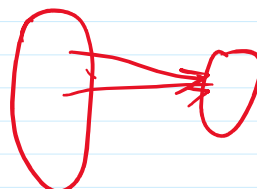


is $A[y] = 1$?

Method 1: hash table with chaining

Pick a hash fn $h: \{0, \dots, U-1\} \rightarrow \{0, \dots, m-1\}$
 for some $m \ll U$.

Store array $A[0 \dots m-1]$
 where $A[i] =$ list of all $x \in S$
 with $h(x) = i$



e.g. $h(x) = x \bmod m$
 $m=5$

query(y): search the list $A[h(y)]$

If the input is rand. unif. distributed
"on average", each bin has $\sim \frac{n}{m}$ elements

Set $m \approx n \Rightarrow O(n)$ space, $O(1)$ average query time
 $O(n)$ preproc time

Method 2 ... With "universal family [Carter, Wegman '77] of hash functions"

idea - pick a random hash fn!
from a family

Ex Assume U is prime.

Pick rand $a \in \{1, \dots, U-1\}$
rand $b \in \{0, \dots, U-1\}$

Define $h_{ab} : \{0, \dots, U-1\} \rightarrow \{0, \dots, m-1\}$

$$h_{ab}(x) = \left(\underbrace{(ax + b) \bmod U}_{\text{red wavy line}} \right) \bmod m$$

Property For any $x, y \in \{0, \dots, U-1\}$ ($x \neq y$).

called universality $\rightarrow$

$$\Pr_{a,b} [h_{ab}(x) = h_{ab}(y)] \leq O\left(\frac{1}{m}\right).$$

say x and y collide

More strongly, for any $x, y \in \{0, \dots, U-1\}$ ($x \neq y$)
& $i, j \in \{0, \dots, m-1\}$,

and $i \neq j \Rightarrow \Pr_{a,b} [h_{ab}(x) = i \text{ and } h_{ab}(y) = j] < O\left(\frac{1}{m^2}\right)$

$$\& i, j \in \{0, \dots, m-1\},$$

called
2-universality $\rightarrow$

$$\Pr_{a,b} [h_{a,b}(x)=i \& h_{a,b}(y)=j] \leq O\left(\frac{1}{m^2}\right)$$

pf: For $s, t \in \{0, \dots, U-1\}$,

$$\Pr_{a,b} \left[\begin{array}{l} (ax+b) \bmod U = s \text{ and} \\ (ay+b) \bmod U = t \end{array} \right]$$

$$= \Pr_{a,b} \left[\begin{array}{l} ax+b \equiv s \pmod{U} \\ ay+b \equiv t \pmod{U} \end{array} \right]$$

$$= \frac{1}{U(U-1)}$$

2x2 linear system
has uniq ans.

$$\begin{cases} a = \frac{s-t}{x-y} \pmod{U} \\ b = s - ax \end{cases}$$

For $i, j \in \{0, \dots, m-1\}$,

$$\Pr_{a,b} [h_{a,b}(x)=i \& h_{a,b}(y)=j]$$

$$= \sum_{\substack{s, t \in \{0, \dots, U-1\} \\ s \bmod m = i \\ t \bmod m = j}} \Pr_{a,b} [\dots]$$

$$\leq \frac{U}{m} \cdot \frac{U}{m} \cdot \frac{1}{U(U-1)}$$

$$\leq \frac{1}{m^2}.$$

□

Query time analysis:

for a fixed query value y ,

$$E \left[\begin{array}{l} \# \text{ collisions} \\ \text{with } y \end{array} \right] = E_{a,b} \left[\# \{x \in S - \{y\} : h_{a,b}(x) = h_{a,b}(y)\} \right]$$

$$E[\text{... with } y] = E_{a,b}[\dots]$$

$$= E_{a,b}\left[\sum_{x \in S - \{y\}} I_x\right]$$

where $I_x = \begin{cases} 1 & \text{if } h_{a,b}(x) = h_{a,b}(y) \\ 0 & \text{else} \end{cases}$

$$= \sum_{x \in S - \{y\}} E_{a,b}[I_x]$$

linearity of expectation

$$= \sum_{x \in S - \{y\}} 1 \cdot \Pr[h_{a,b}(x) = h_{a,b}(y)]$$

$$\leq \sum_{x \in S - \{y\}} O\left(\frac{1}{m}\right) = \underline{O\left(\frac{n}{m}\right)}$$

Set $m \approx n \Rightarrow$

- $O(1)$ expected query time
- $O(n)$ space, preproc time
- $(O(1))$ expected insert/delete time

Q: can we get good worst case query time?

assume static

Alternative query analysis:

$$E[\text{total \# colliding pairs } (x,y) \in S \times S]$$

$$= E\left[\sum_{y \in S} (\text{\# collisions with } y)\right]$$

$$= \sum_{y \in S} E[\text{\# collisions with } y]$$

$$= O\left(n \cdot \frac{n}{m}\right)$$

$$= O\left(n \cdot \frac{n}{m}\right)$$

$$\text{Set } m = cn^2 \Rightarrow E(\text{total \# colliding pairs}) \leq O(1/c)$$

$$\Pr(\text{total \# colliding pairs} \geq 1) \leq \frac{1/c}{1} = \frac{1}{c}$$

Markov's ineq.

$$\Pr(\text{perfect hash}) \geq 1 - \frac{1}{c}$$

$$\Rightarrow \begin{cases} O(n^2) \text{ space} \\ O(1) \text{ query time worst-case} \\ O(n^2) \text{ preproc time } \underline{\text{expected}} \end{cases} \quad (\text{repeat w. success})$$