6. Universal Hashing

2018年10月16日 17:37

U: the universe, all the possible things to be stored
$h: U \to \{0, \dots, m-1\}$ a hash function that hashes need-to-store elements into an array
T: the array of linked lists that stores elements
$S \subseteq U$: wanted keys
if two elements are hashed to the same index, then it is stored as the head of linked list.
Issue: $\forall h. \exists S \subseteq U. \exists i \in \{0, \dots, m-1\}. h(S) = i \wedge |S| \geq \frac{|U|}{m}$

One approach: Average Case Analysis
The input x is sampled from a distribution D, D models of 'nature'.
$t_A(x)$ is a random variable, average runtime = $E\left(t_A(x)\right) = \sum_{t(x) \in S} tP(t_A(x) = t)$
given the assumption that $t_A(S)$ is always finite

Given the Simple uniform hashing assumptions, then it will be constant time.
Assuming D and h are such that $S \sim D$ (random sample of D) $S = \{x_1, \dots, x_n\}$
then $\forall i \in \{1, \dots, n\}. j \in \{0, \dots m-1\}\ AND\ \forall x_i, x_j \in S. i \neq j. h(x_i)$ is independent of $h\left(x_j\right)$

Alternative approach: Randomized Algorithms
alg that makes random choices i.e. $t_A(x)$ is a random variable for each x,
Worst Case Expected Running time $T(n) = \max\left\{ E\left(t_A(x)\right) \mid size(x) = n \right\}$
Intuition: if h has uniformly random funciton, then $E(\#collision)$ would be small

universal a set $\mathcal{H} = \left\{h: U \to \{0, \dots, m-1\}\right\}. \forall x \neq y \in U. h \sim \mathcal{H}.\ P\left(h(x) = h(y)\right) = \frac{\left|\left\{h \in \mathcal{H} | h(x) = h(y)\right\}\right|}{|\mathcal{H}|} \leq \frac{1}{m}$
Goal: find a family of function that takes constant time to properly hash

Theorem if $h \sim \mathcal{H}, S \subseteq U$ is hashed into T using h, then $\forall x \notin S. E\left(n_{h(x)}\right) = \frac{n}{m}. \forall x \in S. E\left(n_{h(x)}\right) < 1 + \frac{n}{m}\ where\ n_i: |T[i]|$

Define the indictor function $I_{xy} = 1\left(h(x) = h(y)\right), E\left(I_{xy}\right) = P\left(I_{xy} = 1\right) = P\left(h(x) = h(y)\right)$

$$n_{h(x)} = \sum_{y \in S} I_{xy}, E\left(n_{h(x)}\right) = E\left(\sum_{y \in S} I_{xy}\right) = \sum_{y \in S} E\left(I_{xy}\right) = \sum_{y \in S} P\left(h(x) = h(y)\right)$$

Consider the cases $x \notin S. \sum_{y \in S} P\left(h(x) = h(y)\right) = m(m\ terms) \times \frac{1}{n}(by\ difinition\ of\ universal)$
if $x \in S, \sum_{y \in S} P\left(h(x) = h(y)\right) = 1(x \in S) + \frac{(n-1)}{m} < 1 + \frac{n}{m}$

Corollary $h \in O(1) \Rightarrow insert \in O(1), search \in O\left(n_{h(x)}\right) = O\left(1 + \frac{n}{m}\right)$
Example $U = \{0, \dots, p-1\}. p \in \mathcal{P}$ =set of prime number
$\mathcal{H} = \left\{h_{ab}: a \in \{1, \dots, p-1\}, b \in \{0, \dots, p-1\}\right\}, h_{ab}(x) := \left((ax+b)(\mathrm{mod}\ p)\right) \bmod m$
Show $\mathcal{H}$ is universal