2012-12-05

DSEA 4

$\mathbb{Z}_p : \exists a^{-1} \; \forall a \neq 0 \qquad p = 7$

$3^{-1} = 5 \qquad 3 \cdot 5 \bmod 7 = 1$

$\Rightarrow$ erweiterter Euklidischer Algo.

2012-12-10

DSEA 1

Idealisierte Annahme: Die n Elemente werden unabh. und gleichverteilt auf die m Tabellenplätze abgebildet.

$h : U \to \{0, \dots, m-1\}$

Frage: Wie hoch ist die mittlere Suchzeit für ein zufälliges Element aus der Tabelle?

Sei $k_1, k_2, \dots, k_n$ die Einfügereihenfolge der n Schlüssel in die Tabelle.

$$X_{ij} = \begin{cases} 1 & \text{falls } h(k_i) = h(k_j) \\ 0 & \text{sonst} \end{cases}$$

$$E(T) = \sum_{i=1}^{n} \frac{1}{n} E\Big(\underbrace{1}_{\text{Hashfkt. auswerten}} + \underbrace{\sum_{j=i+1}^{n} X_{ij}}_{\text{Anzahl der Elemente, die nach } k_i \text{ kommen und mit } k_i \text{ kollidieren}}\Big)$$

$$E(X_{ij}) = 0 \cdot P(X_{ij} = 0) + 1 \cdot P(X_{ij} = 1) = P(X_{ij} = 1) = \frac{1}{m}.$$

$$\Rightarrow E(T) = \sum_{i=1}^{n} \frac{1}{n} \Big(1 + \sum_{j=i+1}^{n} E(X_{ij})\Big) = 1 + \frac{1}{n} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{1}{m} = 1 + \frac{1}{n \cdot m} \cdot \binom{n}{2}$$

$$= 1 + \frac{n(n-1)}{2nm} \leq 1 + \frac{\alpha}{2} \quad , \alpha = \text{Belegungsfaktor}.$$

2012-12-10

DSEA 2

Universelles Hashing

$\mathcal{H}$ Klasse von Hashfktn.

z.B. $h(k) = (ak+b) \bmod p \bmod m$ ~~[illegible]~~

$1 \leq a < p$

$0 \leq b < p$

$m < p$ prim

<u>Def.</u>: $\mathcal{H}$ heißt universell :$\Leftrightarrow$ $\forall k \neq l \in U: \left|\{h \in \mathcal{H} \mid h(k) = h(l)\}\right| \leq \frac{|\mathcal{H}|}{m}$

d.h. $P(h(k) = h(l)) \leq \frac{1}{m}$ für eine zufällig gewählte Hashfktn. $h \in \mathcal{H}$.

Sei n_i die erwartete Länge der Kollisionsliste bei Feldeintrag i.

Y_k = Länge der Kollisionsliste für den Schlüssel k.

$$\Rightarrow n_{h(k)} = \begin{cases} E(Y_k), & \text{für } k \notin T \\ E(Y_k) + 1 & \text{für } k \in T \end{cases}$$

$Y_k = \sum_{l \in T \setminus \{k\}} X_{kl}$ T: Tabelle

$$X_{kl} = \begin{cases} 1 & \text{falls } h(k) = h(l) \\ 0 & \text{sonst} \end{cases}$$

$$E\Big(\sum_{l \in T \setminus \{k\}} X_{kl}\Big) = \sum_{l \in T \setminus \{k\}} E(X_{kl}) \leq \sum_{l \in T \setminus \{k\}} \frac{1}{m}$$

1\. Fall: $k \in T \Rightarrow n_{h(k)} = \frac{n-1}{m} + 1 \leq 1 + \alpha$

2\. Fall: $k \notin T \Rightarrow n_{h(k)} = \frac{n}{m} = \alpha$

$\Rightarrow$ Erwartete Länge aller Kollisionslisten ist in $O(1+\alpha)$.

2012-12-10

DSGA.3

Vereinfachte Annahme:

$$\mathcal{H} = \left\{ \begin{array}{l} h(k) = (a \cdot k) \bmod p \bmod m \\ 1 \le a < p,\ m < p \text{ Primzahl} \end{array} \right\}$$ (d.h. ohne $+b$, kann später hinzugefügt werden, ändert nichts)

$|\mathcal{H}| = p-1.$

Wähle $k \neq l \in U$ fest, $k, l < p$

$h(k) = h(l) \Leftrightarrow \quad ak \bmod p \bmod m = al \bmod p \bmod m$

a.B.d.A. $ak \bmod p > al \bmod p$.

$x_1 \equiv x_2 \bmod m \quad \Leftrightarrow \quad x_1 = x_2 + i \cdot m$

$x_1 \geq x_2$

$\Leftrightarrow \quad ak \bmod p = al \bmod p + i \cdot m \qquad 0 < i \leq \left\lfloor \frac{p-1}{m} \right\rfloor$

$i \neq 0$: Ann: $ak \equiv al \bmod p$

$a(k-l) \equiv 0 \bmod p$

$\lightning$

$\Leftrightarrow \quad a(k-l) \equiv im \bmod p$

$a \equiv -(k-l)^{-1} \cdot im \bmod p$ $\Rightarrow \forall i\ \exists!\, a$ in $\mathbb{Z}_p$ so dass $h(k) = h(l)$

$|\{h \in \mathcal{H} \mid h(k) = h(l)\}| \leq \frac{p-1}{m} = \frac{|\mathcal{H}|}{m}.$

ÜB: andere Hashfkt.

2012-12-10

DSEA 4

Perfektes Hashing

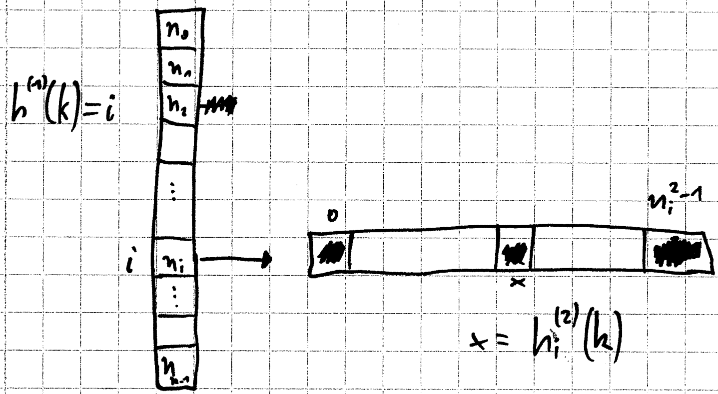
Bisher: Suche entspricht im Mittel dem Belegungsfaktor. Wollen: garantierten konstanten Zugriff.

Statisches Szenario (kein Einfügen/Löschen), n Elemente a priori bekannt.

Frage: Wie kann man garantieren, dass Suchzeit in $O(1)$ liegt?

Idee: Zweistufige Hashtabelle.

$m := n$, n_i: Anzahl der Elemente in den Listen



$P(\exists$ Kollision zwischen zwei Elementen aus n_i Elementen, bei zufälliger Wahl von $h_i^{(2)}$ und Tabellengröße $n_i^2)$

↑ universelle H-Fkt.

$$\leq P(\exists \text{ Kollision zwischen einem festen Paar}) \cdot \binom{n_i}{2}$$

$$\overset{\text{universell}}{\leq} \frac{1}{n_i^2} \cdot \binom{n_i}{2} = \frac{n_i(n_i-1)}{2n_i^2} \leq \frac{1}{2}.$$

Erwarteter Platzbedarf für Sekundärtabellen

$$E(n_i^2) = E\left(n_i + 2\frac{n_i(n_i-1)}{2}\right) = E\left(n_i + 2\binom{n_i}{2}\right)$$

$$E\left(\sum_{i=0}^{n-1} n_i^2\right) = E\left(\sum_{i=0}^{n-1} n_i\right) + 2 \cdot \underbrace{E\left(\sum_{i=0}^{n-1} \binom{n_i}{2}\right)}_{\text{Summe der Kollisionen}}$$

$$= n + 2 \cdot E\left(\sum_{i=0}^{n-1} \sum_{\substack{k \neq l \\ \in T_i}} X_{kl}\right) = n + 2 \cdot \sum_{i=0}^{n-1} \sum_{k \neq l} E(X_{kl})$$

$$= n + 2 \sum_{i=0}^{n-1} \binom{n_i}{2} \frac{1}{n_i} = n + 2n\frac{1}{2} \leq 2n.$$

→ linear.