Script generated by TTT

Title: TÃ¤ubig: GAD (16.05.2013)

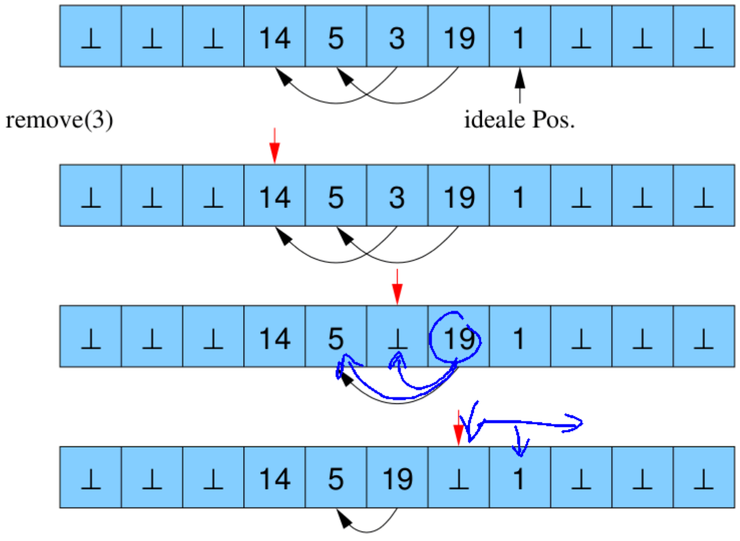
Date: Thu May 16 12:02:53 CEST 2013

Duration: 48:36 min

Pages: 14

Hashing with Linear Probing

Löschen / Aufrechterhaltung der Invariante



Übersicht

1|17|9| ⊥

5 Hashing
- Hashtabellen
- Hashing with Chaining
- Universelles Hashing
- Hashing with Linear Probing
- Anpassung der Tabellengröße
- Perfektes Hashing
- Diskussion / Alternativen

Dynamisches Wörterbuch

Problem: Hashtabelle ist zu groß oder zu klein
(sollte nur um konstanten Faktor von Anzahl der Elemente abweichen)

Lösung: Reallokation

- Wähle geeignete Tabellengröße
- Wähle neue Hashfunktion
- Übertrage Elemente auf die neue Tabelle

Dynamisches Wörterbuch

Problem: Tabellengröße m sollte prim sein
(für eine gute Verteilung der Schlüssel)

Lösung:

- Für jedes k gibt es eine Primzahl in $\left[k^3, (k+1)^3\right]$
- Jede Zahl $z \leq (k+1)^3$, die nicht prim ist, muss einen Teiler $t \leq \sqrt{(k+1)^3} = (k+1)^{3/2}$ haben.
- Für eine gewünschte ungefähre Tabellengröße m' (evt. nicht prim) bestimme k so, dass $k^3 \leq m' \leq (k+1)^3$

Dynamisches Wörterbuch

$(a+b)^n = \binom{n}{0} a^n b^0 + \binom{n}{1} a^{n-1} b^1 + \dots$

Problem: Tabellengröße m sollte prim sein
(für eine gute Verteilung der Schlüssel)

Lösung:

- Für jedes k gibt es eine Primzahl in $\left[k^3, (k+1)^3\right]$
- Jede Zahl $z \leq (k+1)^3$, die nicht prim ist, muss einen Teiler $t \leq \sqrt{(k+1)^3} = (k+1)^{3/2}$ haben.
- Für eine gewünschte ungefähre Tabellengröße m' (evt. nicht prim) bestimme k so, dass $k^3 \leq m' \leq (k+1)^3$
- Größe des Intervalls: $(k+1)^3 - k^3 + 1 = (\cancel{k^3} + 3k^2 + 3k + 1) - \cancel{k^3} + 1 = 3k^2 + 3k + 2$
- Für jede Zahl $j = 2, \ldots, (k+1)^{3/2}$: streiche die Vielfachen von j in $[k^3, (k+1)^3]$ $O(k^2)$
- Für jedes j kostet das Zeit $\left((k+1)^3 - k^3 + 1\right)/j \in O(k^2/j)$

Dynamisches Wörterbuch

- Hilfsmittel: Wachstum der harmonischen Reihe

$$\ln n \quad \leq \quad H_n = \sum_{i=1}^{n} \frac{1}{i} \quad \leq \quad 1 + \ln n$$

- insgesamt:

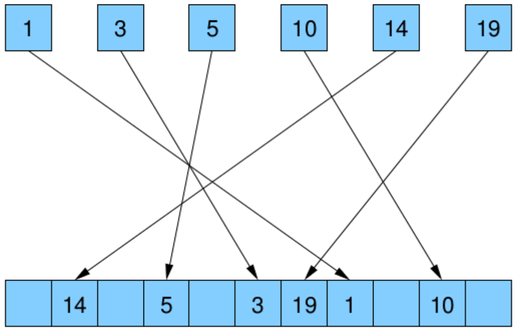
$$\begin{aligned} \sum_{2 \leq j \leq (k+1)^{3/2}} O\left(\frac{k^2}{j}\right) &\leq k^2 \sum_{2 \leq j \leq (k+1)^{3/2}} O\left(\frac{1}{j}\right) \\ &\in k^2 \cdot O\left(\ln\left((k+1)^{3/2}\right)\right) \\ &\in O(k^2 \ln k) \in o(k^3) \log k \in o(k \\ &\in o(m) \end{aligned}$$

$\Rightarrow$ Kosten zu vernachlässigen im Vergleich zur Initialisierung der Tabelle der Größe m (denn m ist kubisch in k)

Übersicht

Perfektes Hashing für statisches Wörterbuch

- bisher: konstante *erwartete* Laufzeit
  falls m im Vergleich zu n genügend groß gewählt wird
  (nicht ausreichend für Real Time Scenario)
- Ziel: konstante Laufzeit im worst case für find()
  durch perfekte Hashtabelle ohne Kollisionen
- Annahme: statische Menge S von n Elementen



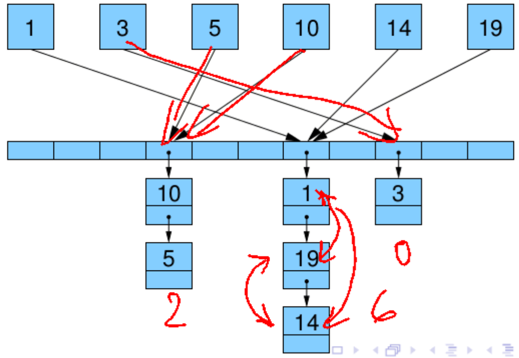
Statisches Wörterbuch

- S: feste Menge von n Elementen mit Schlüsseln k_1 bis k_n
- H_m: c-universelle Familie von Hashfunktionen auf $\{0, \ldots, m-1\}$
  (Hinweis: 2-universelle Familien existieren für alle m)
- $C(h)$ für $h \in H_m$: Anzahl Kollisionen in S für h, d.h.

$$C(h) = \left|\{(x, y) : x, y \in S,\ x \neq y,\ h(x) = h(y)\}\right|$$

Beispiel:

$C(h) = 2 + 6 = 8$



Statisches Wörterbuch

Lemma

Es gilt

$$\mathbb{E}[C(h)] \leq cn(n-1)/m$$

Statisches Wörterbuch

Lemma

Es gilt

$$\mathbb{E}[C(h)] \leq cn(n-1)/m$$

Beweis.

- Definiere $n(n-1)$ Indikator-Zufallsvariablen $X_{ij}(h)$:
  Für $i \neq j$ sei $X_{ij}(h) = 1 \ \Leftrightarrow\ h(k_i) = h(k_j)$.
- Dann ist $C(h) = \sum_{i \neq j} X_{ij}(h)$

$$\mathbb{E}[C] = \mathbb{E}\left[\sum_{i \neq j} X_{ij}\right] = \sum_{i \neq j} \mathbb{E}\left[X_{ij}\right] = \sum_{i \neq j} \Pr\left[X_{ij} = 1\right] \leq n(n-1) \cdot c/m$$

$m \in \Theta(n^2)$

$\Rightarrow$ Für quadratische Tabellengröße ist die erwartete Anzahl von Kollisionen (und damit die erwartete worst-case-Laufzeit für find) eine Konstante. □

Markov-Ungleichung

Satz (Markov-Ungleichung)

Für jede nichtnegative Zufallsvariable X und Konstante k gilt:

$$\Pr[X \geq k \cdot \mathbb{E}[X]] \leq \frac{1}{k}$$

Markov-Ungleichung

Satz (Markov-Ungleichung)

Für jede nichtnegative Zufallsvariable X und Konstante k gilt:

$$\Pr[X \geq k \cdot \mathbb{E}[X]] \leq \frac{1}{k}$$

Beweis.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{z \in \mathbb{R}} z \cdot \Pr[X = z] \\
&\geq \sum_{z \geq k \cdot \mathbb{E}[X]} z \cdot \Pr[X = z] \geq \sum_{z \geq k \cdot \mathbb{E}[X]} k \cdot \mathbb{E}[X] \cdot \Pr[X = z] \\
&\geq k \cdot \mathbb{E}[X] \cdot \Pr[X \geq k \cdot \mathbb{E}[X]]
\end{aligned}$$

$\frac{1}{k} \geq \Pr[\ldots]$

$\square$

Statisches Wörterbuch

Lemma

Für mindestens die Hälfte der Funktionen $h \in H_m$ gilt:

$$C(h) \leq 2cn(n-1)/m$$