Prof. Dr. Heiner Klocke
Algorithmik SS 2006

Übung: Hashtechniken mit offener Adressierung

Sei k der zu speichernde Schlüssel, sei m die Größe der Hashtabelle.

Linear Probing ist definiert durch die Hashfunktion

$$h(k,i) = (h'(k) + i) \bmod m, \quad i = 0, 1, ..., m-1$$

Quadratic Probing ist definiert durch

$$h(k,i) = (h'(k) + c_1 i + c_2 i^2) \bmod m, \quad i = 0, 1, ..., m-1$$

h' ist die primäre Hashfunktion; c_1 und c_2 sind Konstanten ungleich 0.

Double Hashing ist definiert durch

$$h(k,i) = (h_1(k) + i * h_2(k)) \bmod m, \quad i = 0, 1, ..., m-1$$

Aufgabe 1

Wie lauten die Sondierungsfolgen für Linear Probing, Quadratic Probing und Double Hashing? Erklären Sie das Problem der primären und sekundären Häufung.

Aufgabe 2

Folgende Schlüssel 10, 22, 31, 4, 15, 28, 17, 88 und 59 sollen in eine Hashtabelle der Größe m = 11 mit offener Adressierung eingefügt werden. Benutzen Sie als primäre Hashfunktion **h(k) = k mod m**.

Wie lauten die Ergebnisse für

(a) Linear Probing

(b) Quadratic Probing mit $c_1 = 1$, $c_2 = 3$

(c) Double Hashing mit $\mathbf{h_1(k) = k \bmod m}$ und $\mathbf{h_2(k) = 1 + (k \bmod (m-1))}$

Aufgabe 3

Im folgenden sehen Sie die theoretischen Analysen für die Performance der Hashverfahren *Linear Probing*, *Quadratic Probing* und *uniformes Hashing*.

a) Was versteht man bei Hashverfahren unter Performance?

b) Interpretieren Sie die Tabellen. Vergleichen Sie die Tabellen, und erklären Sie die Unterschiede.

c) Welche Experimente würden Sie durchführen, um die Daten in den Tabellen experimentell zu überprüfen?

Linear Probing

$$C_{n'} \approx \frac{1}{2}\left(1 + \frac{1}{(1-a)^2}\right) \qquad \text{(erfolglose Suche)}$$

$$C_n \approx \frac{1}{2}\left(1 + \frac{1}{1-a}\right) \qquad \text{(erfolgreiche Suche)}$$

a	erfolglos	erfolgreich
10%	1.11	1.0
30%	1.52	1.21
50%	2.5	1.5
70%	6.0	2.17
90%	50.5	5.5
95%	200.5	10.5
98%	1250.5	25.5

Was bedeuten die Zahlen in dieser Tabelle?

Quadratic Probing

Erfolgreiche Suche: $C_n \approx 1 + \ln\left(\frac{1}{1-a}\right) - \frac{a}{2}$

Erfolglose Suche: $C_{n'} \approx \frac{1}{1-a} - a + \ln\left(\frac{1}{1-a}\right)$

a	erfolglos	erfolgreich
10%	1.11	1.0
30%	1.48	1.20
50%	2.19	1.44
70%	3.84	1.85
90%	11.4	2.85
95%	22.04	3.52
98%	53	4.4

Uniformes Hashing

Erfolgreiche Suche $$C_n \approx \frac{1}{a} \cdot \ln\left(\frac{1}{1-a}\right)$$

Erfolglose Suche $$C_n' \approx \frac{m}{m-n} = \frac{1}{1-a}$$

a	erfolglos	erfolgreich
10%	1.1	1.0
30%	1.42	1.18
50%	2.0	1.38
70%	3.34	1.72
90%	10	2.56
95%	20	3.15
98%	50	4.0

Aufgabe 4

a) Was versteht man unter einer *uniformen* und was unter einer *zufälligen* Verteilung von Schlüsseln auf die Slots einer Hashtabelle?

b) Sei $m = 2^8 = 256$ die Größe der Hashtabelle, und sei die Hashfunktion definiert durch $h(k) = k \bmod 256$, wobei die Schlüssel k jeweils aus drei 8-Bit ASCII-Zeichen bestehen. Jeder Schlüssel k kann also durch eine 24-Bit Integerzahl repräsentiert werden: $C_3C_2C_1$

Welches Problem tritt bei der Hashfunktion $h(k) = k \bmod 256$ auf? Konstruieren Sie ein Beispiel.