

Aufgabe 4.1 4 Punkte.

Auf dem Mars dauert ein Jahr 687 Tage, d.h. ein Marsianer hat an einem von 687 Tagen Geburtstag.

- Wie viele Marsianer müssen zusammenkommen damit die Wahrscheinlichkeit, dass zwei am gleichen Tag Geburtstag haben größer als $\frac{2}{3}$ ist.
- Bestimmen Sie eine allgemeine Formel, die angibt, bei wie vielen Einträgen in einer Hashtabelle die Kollisionswahrscheinlichkeit $> \frac{2}{3}$ ist.

Aufgabe 4.2 1+2+2=5 Punkte.

Es sei die Schlüsselmenge $S = \{92, 19, 83, 37, 16, 57, 61\}$ und eine Hashtabelle der Größe $m = 11$ gegeben. Illustrieren Sie das Einfügen der Schlüssel mit den folgenden Hashverfahren. Geben Sie zusätzlich die Anzahl der Kollisionen an.

- Hashing mit Verkettung: $h(s) = s \bmod m$
- Hashing mit linearem Sondieren: $h(s, i) = (h_1(s) + i) \bmod m$ mit $h_1(s) = s \bmod m$.
- Double Hashing: $h(s, i) = (h_1(s) + ih_2(s)) \bmod m$,
mit $h_1(s) = s \bmod m$ und $h_2(s) = 1 + (s - 1 \bmod (m - 1))$

Hinweis: In (b) und (c) ist s der Schlüssel und i die Anzahl der bisherigen Kollisionen für den Schlüssel s . Bei 0 Kollisionen wird $h(s, 0)$ ausgerechnet, wenn dieser kollidiert, dann wird $h(s, 1)$ versucht, usw.

Aufgabe 4.3 2 Punkte.

Betrachten Sie *Hashing mit Verkettung*: Sei U eine Menge von Schlüsseln, die potentiell gehasht werden können (U sei das Universum), $m \geq 2$ sei die Hashtabellengröße. Zeigen Sie, dass für jede Hashfunktion h eine Menge $S \subset U$ von n Schlüsseln existiert ($n = |S|$), die auf die gleiche Position in der Hashtabelle abgebildet werden, wenn $|U| > m(n - 1)$ gilt.

Aufgabe 4.4 2 Punkte.

Bloomfilter erlauben es eigentlich nicht, dass eingeprägte Werte wieder gelöscht werden. Student Oberhuber implementiert trotzdem eine Löschoperation. Diese setzt bei der Löschung von x alle Einträge $A[y_i]$ mit $y_i = h_i(x)$ auf *false*. ($\{h_i \mid 1 \leq i \leq k\}$ ist die Menge der verwendeten Hashfunktionen).

Welche Probleme können durch diese Löschoperation auftreten?

Aufgabe 4.5 3 Punkte.

Gegeben sei ein Bloomfilter mit $m = 1000$. Eine Schlüsselmenge S mit $|S| = 100$ wird in das Array „eingeprägt“. Ermitteln Sie diejenige Anzahl $k \in \{2, \dots, 12\}$ an Hashfunktionen, die die Wahrscheinlichkeit für false positives minimiert. Geben Sie auch die auf 4 Nachkommastellen gerundete Wahrscheinlichkeit an.

Aufgabe 4.6 *4 Punkte.*

Lösen Sie die Programmier-Aufgabe "Datenbankhack".

Geben Sie Ihren Domjudge-Teamnamen bei Ihrer Abgabe an, damit Ihnen Ihre Lösung zugeordnet werden kann.