

Big Data Analytics in Theorie und Praxis

Sommersemester 2016

Theorie-Übungsblatt 4

Zu bearbeiten bis 12. Juli 2016

Aufgabe 1: (25 Punkte)

Seien $a, z, p \in \mathbb{N}$ mit $a < p$ und $z < p$. Beschreiben Sie einen Algorithmus mit möglichst geringer Laufzeit, der $z^a \bmod p$ mit $\mathcal{O}(\log p)$ Speicherbits berechnet. Weisen Sie nach, dass Ihr Algorithmus tatsächlich $z^a \bmod p$ korrekt bestimmt und höchstens $\mathcal{O}(\log p)$ Speicherbits benutzt. Bestimmen Sie die Laufzeit, die Ihr Algorithmus benötigt.

Aufgabe 2: (35 Punkte)

Betrachten Sie die in der Vorlesung definierten Familien H_1 , H_2 und H_{s2} von Hash-Funktionen.

(a) Zeigen Sie, dass H_1 streng 2-universell ist.

(b) Zeigen Sie, dass H_2 2-universell ist.

Ist H_2 auch *streng* 2-universell? Geben Sie konkrete Zahlen m , M und p an, für die die Familie H_2 tatsächlich streng 2-universell ist. Und geben Sie (andere) Parameter m , M und p an, für die Sie nachweisen können, dass die Familie H_2 *nicht streng* 2-universell ist.

(c) Zeigen Sie, dass H_{s2} streng 2-universell ist.

Aufgabe 3: (40 Punkte)

In der Vorlesung wurde skizziert, wie man die Datenstruktur *Count-Min-Sketch* dazu verwenden kann, *Bereichsanfragen* zu beantworten, indem man $\mathcal{O}(\log m)$ viele Count-Min-Sketches verwendet und den angefragten Bereich $[\ell, \dots, r]$ in $\mathcal{O}(\log m)$ viele dyadische Bereiche zerlegt. Arbeiten Sie die Details dazu aus! Geben Sie dabei auch an

- wieviel Speicherplatz die gesamte Datenstruktur verbraucht,
- wie lange es beim Lesen eines Datenstromelements (u, z) dauert, um die Datenstruktur zu aktualisieren
- wie bei Eingabe von Zahlen ℓ und r die zugehörige Bereichsanfrage $f_{[\ell, \dots, r]}$ beantwortet wird und
- welche Garantie (formuliert unter Verwendung von ε und δ) Sie über die Güte des ausgegebenen Schätzwerts geben können.