# Proseminar WS 2016/2017

## Bloom-Filter

**Gegeben:** Eine Menge $M$ natürlicher Zahlen und $n \in \mathbb{N}$.
**Gesucht:** Prüfe in konstanter Zeit, ob $n \in M$ gilt.

**Naive Methode**: Ist $N$ die größte Zahl in $M$, so erzeugen wir ein Bitarray $A_M$ der Größe $N$. Für $m = 1, \ldots, N$ gilt

$$A_M[m] = \begin{cases} 1, & \text{wenn } m \in M \\ 0, & \text{sonst.} \end{cases}$$

Es gilt $n \in M \Longleftrightarrow A[n] = 1$ gilt.

**Bemerkung:** Je größer $N$ wird, desto mehr Speicherplatz benötigen wir mit der naiven Methode. Daher benötigen wir eine effizientere Datenstruktur.

**Bloom-Filter**: Seien $\phi_1, \ldots, \phi_k$, $k$-verschiedene Hashfunktionen (mit gewissen Eigenschaften) und $A_M$ ein Bitarray der Größe $l \in \mathbb{N}$, $l \approx |M|^2$. Für alle $m \in M$ und $j = 1, \ldots, l$ gilt

$$A_M[j] = 1 \Longleftrightarrow \exists m \in M \text{ mit } \phi_i(m) = j.$$

Die Frage ob $n \in M$ ist, können wir mit einer gewissen Wahrscheinlichkeit ebenfalls in konstanter Zeit beantworten. Ist jedoch $A_M[\phi_i(n)] = 0$ für ein $i \in \{1, \ldots, k\}$, so können wir sicher sein, dass $n$ kein Element aus $M$ ist.

**Fazit:** Ist $N \gg |M|$, so hat die naive Methode erhebliche Nachteile gegenüber dem Bloom-Filter, der unabhängig von $N$ ist.

In Ihrer Ausarbeitung geht es darum, die nötige Theorie zu entwickeln, um den Bloom-Filter zu beschreiben.

Empfohlene Literatur:


- U. Schöning. Skript zu Algorithmen und Datenstrukturen
- M. Mitzernmacher, E. Upfal. Propability and Computing.