Heute gebe ich euch einen kurzen Überblick über eine ganz besondere, nämlich probabilistische Datenstruktur und zwar den Bloom Filter. Was das ist und was ihn auszeichnet seht ihr jetzt.

* Erfunden 1970 durch Burton Bloom
* erhält seitdem viel Aufmerksamkeit sowohl in Literatur als auch Praxis
* Mit ihm schnell überprüfen ob Element Teil einer bestimmten Menge Elemente ist
* Dafür wird Element mittels Hash Fkt auf Position in Array heruntergebrochen
* Anfangs alles in Array 0
* Dann jede Position zu 1 bei Insert
* Wenn alle Elemente eingetragen kann man auf gleiche Weise testen
* Wieder Hash bilden und schauen ob überall 1
* Wenn nur eine 0, kann Element nicht dabei gewesen sein
* Wenn alles 1 wird „Ja“ ausgegeben, aber besteht auch Chance, dass nicht drin war
* Positionen können ja auch durch andere Elemente besetzt wurden sein
* Somit False positive Rate 🡪 probabilistisch
* Dadurch eben platzsparend, weil Elemente ja selbst nicht speichert
* Nur 0 und 1
* So viele Einsatzmöglichkeiten wie bei Rechtschreibprüfung
* Alle Wörter der Sprache in BF und dann eingegebene Elemente geprüft ob Element davon
* Suchmaschinen 🡪 Dokumente werden in BF eingetragen und ähnliche Inhalte = ähnliche verwendete Worte 🡪 verwandte Inhalte finden
* Web Caches prüfen ob bestimmte Seite in Cache
* Sicherheit zB Intrusion Detection, da schnell Schadstrings erkennt