BLOOM FILTERS:

GROUP NAME: AMORTIZED ARRAY

GROUP MEMBER

- Fatima Nasir
- Sahil Suresh
- Hammad Ahmed
- Shayan Sharif
- Gulzar

WHAT ARE BLOOM FILTERS

- Bloom filter tells whether an element is already the part of the larger set or not.
- We have studied hash tables and we have seen how we solve collisions through chain linking and open addressing. These collision problems also occur in bloom filters but they are acceptable to a certain point.

WHAT WE NEED?

- Bloom filters uses fixed size arrays, the array is a bit array.
- One or multiple hash functions.

FALSE POSITIVES

- False positives are instances when the search returns true even though the data is not exactly present in the data structure.

HOW TO TACKLE FALSE POSITIVES?

- The bigger the size of array the less chances of collision of bits.
- More hash functions also lowers risk of collision as there will be more bits to represent in the memory. Also more efficient hash functions can be used to ensure the reduction of false positives.

OPTIMAL HASH FUNCTIONS:

So the number of optimal hash functions can be figured out using :

$$k = \frac{m}{n} \ln 2.$$

THE **PROBABILITY** THAT A BIT 'J' IS NOT SET TO 1 IS

(1–1/n)^n

FALSE NEGATIVES

- However, there is no possibility of false negative in bloom filters; if a value is not present in the data structure it is 100% not there.

APPLICATIONS

- Weak password detection
- Safe browsing in Google Chrome
- Ethereum (cryptocurrency) uses Bloom filters for quickly finding logs on the Ethereum blockchain.

Some simple applications can be,

- The most common use for bloom filters is probably testing to see if an element exists on disk before performing any i/o.
- Unique counts
- Facebook uses bloom filters for typeahead search.

WHY BLOOM FILTERS

Why not use Database and SQL?

- Whereas,

Bloom filter finds an element in constant time in a large set O(K) where K is the number of hash functions used. It looks at the binary array first to tell whether an element exist or not.

PROBLEM STATEMENT

The problem we are working on is related to “Book Reads”

Our program, will keep an account of the books read. It manages the list of read books and tells similar books.

It will be similar to a website called “GoodReads” that allows to track the books he/she has read. One can also Look at the details f book like author, rating, genre and similar books

Our mission is to help people find and share books they love

TASK ALLOCATION

- Most of the front end work will be done by Shayan and Gulzar. While Sahil, Fatima and Hammad will be working on the back end mostly.
- Implementation of bloom filter → Fatima and Sahil
- Dataset handling and implementing hash functions → Hammad and Shayan.

- Once the implementation is started all work will be distributed equally among all the members.

RESEARCH

- We have researched on what Bloom-Filters are and how do they work.
- What should be the size of an array.
- How many number of hashing functions one should use..
- How to tackle problem of False Positive.

We have all our researched work in a work file, that we can share.

RESOURCES

- https://www.youtube.com/watch?v=S_ZEQRJjfPo
- https://www.youtube.com/watch?v=U8Ni1yJ8ZS4
- https://www.youtube.com/watch?v=gBygn3cVP80
- https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.20.98
- http://crystal.uta.edu/~mcguigan/cse6350/papers/Bloom.pdf
- http://pages.cs.wisc.edu/~cao/papers/summary-cache/node8.html
- https://www.jasondavies.com/bloomfilter/
- https://en.wikipedia.org/wiki/Bloom_filter#Extensions_and_applications
- https://www.perl.com/pub/2004/04/08/bloom_filters.html/
- https://blog.medium.com/what-are-bloom-filters-1ec2a50c68ff#.xlkqtn1vy
- https://www.quora.com/What-are-the-best-applications-of-Bloom-filters