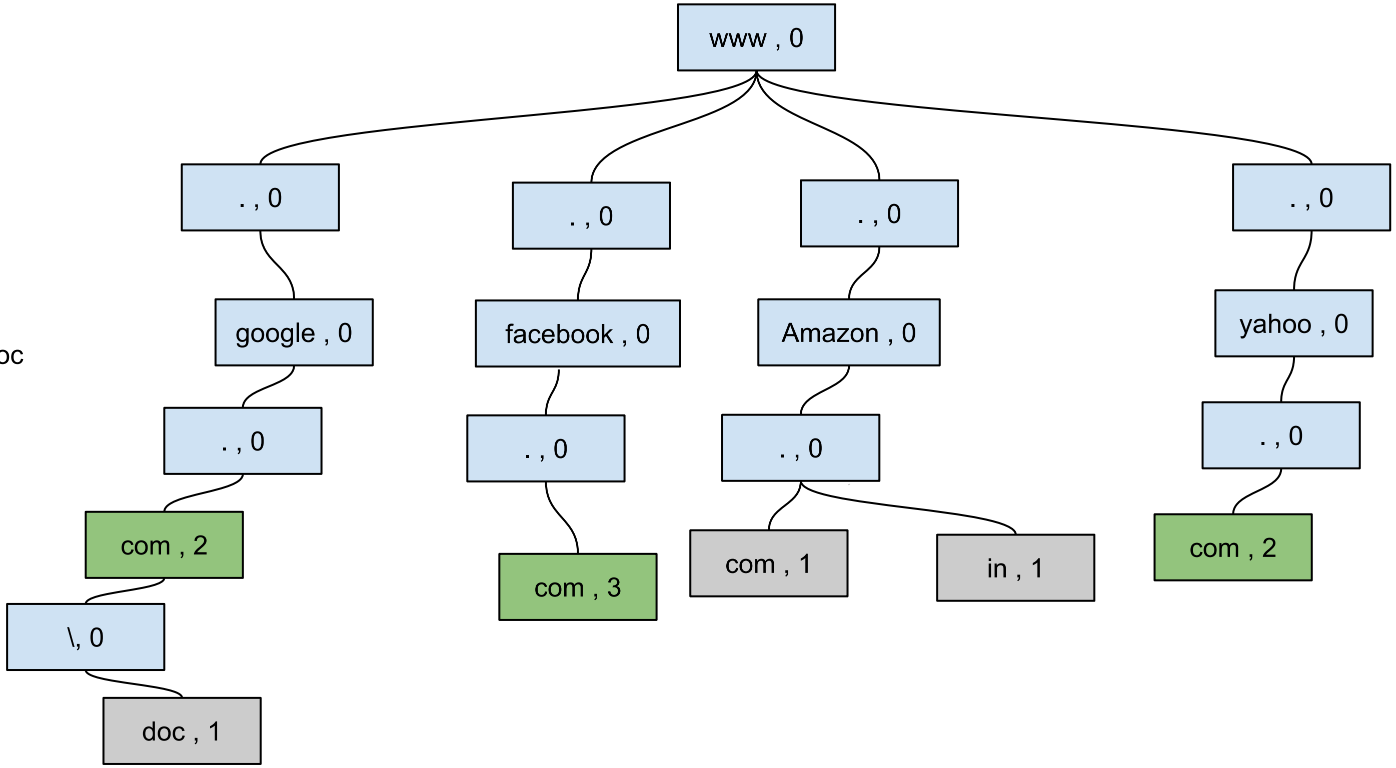
Problem: Identified duplicate URL in very long list of URL say 10b of URL

1. Using hashing of url and use hashmap to identify the duplicate - colloison is the issue
2. Use Bloom filter with 3-4 hash value - again issue in memory and accuracy of 100%
3. Device and conquer - here alternate url matched with next url if duplicate then notify and remove till all duplicate removed.
4. Key-valye in memory cache (REDIS)- user url as key and value if count
5. Use Trie data structure to identify the duplicate url where
    - a. each node in trie contain Char and numeric value which is count of every url end the the Char of this node means duplicate url.
    - b. We can also user word and separator like \ and dot in url instead char in url to keep low depth of tri. It will take more space for trie

List of URL
www.google.com
www.facebook.com
www.amazon.com
www.yahoo.com
www.facebook.com
www.yahoo.com
www.facebook.com
www.google.com
www.amazon.in
www.google.com\doc


www , 0
. , 0
. , 0
. , 0
. , 0
google , 0
facebook , 0
Amazon , 0
yahoo , 0
. , 0
. , 0
. , 0
. , 0
com , 2
com , 3
com , 1
in , 1
com , 2
\, 0
doc , 1