**Problem Statement:**

Generate a random numbers without duplicates within the long values such as 10^7, 10^8 and 10^9. compare the running times of the HashSet and StringSet. Conclude which is the best in both approaches.

**Analysis:**

**1. HashSet:** It is implemented by the hashtable. After generating a random value by using the method StdRandom.uniform(maximum\_value) from java.util.concurrent.ThreadLocalRandom library. Hash code is generated for the value to put into the hash table. All these method are used in the Dedup method for as to eliminate the duplicates.

Memory complexity will be more, for the given N values 10^7 is only executing while 10^8 and 10^9 getting the run time exception “java.lang.OutOfMemoryError: Java heap space”.

**Time complexities :**

Dedup method : O(N)

Add method : O(1)

Contains method : O(1)

**2. StringSet:** In the StringSet, we follow the same process as HashSet as above mentioned. The generated random value is converted into the string by using the valueOf(random\_number) and insert into the data structure. Dedup method complexity is O(N). Here we take the radix as 10 rather than 256, so the memory space will be less.

In **TST**(ternary Search Tree) the space complexity will be more as it depends length of the string. For insert and search methods the time complexity will be depend on the height of the tree. M is length of the string.

Average case Worst case

|  |  |  |
| --- | --- | --- |
| Insert | O(log(M)) | O(M) |
| search | O(log(M)) | O(M) |

In **Trie**, the worst case runtime is dependent on how many strings it contains. M is the longest string and N is the number of strings. So the complexity will be the O(NM). A is the length of string.

**Time complexities :**

Dedup method : O(N)

Insert method: O(AN)

Search method: O(AN)

Even **StringSet** can be implemented through the hashMap. Memory complexity will be more, for the given N values 10^7 is only executing while 10^8 and 10^9 getting the run time exception “java.lang.OutOfMemoryError: Java heap space”.

**Time complexities :**

Dedup method : O(N)

Add method : O(1)

Contains method : O(1)

**Conclusion:**

To decide which is best the approach for this problem is depends on the consideration of time and space complexities. While comparing the two approaches, HashSet has less time complexity and more space complexity, StringSet has more time complexity and less space complexity. In StringSet , if we consider R-way trie over tst.It has less time complexity and more space complexity. As the TST has more time complexity and less space complexity. In R-way trie,there is also possibility that to get the out of memory exception.For searching TST is the best as it take log(N).