**Problem Statement :- 1**

Redo Exercise 3.5.30 using StringSET (see Exercise 5.2.6) instead of HashSET. Compare the running times of the two approaches. Then use Dedup to run the experiments for N = 10^7, 10^8, and 10^9, repeat the experiments for random long values and discuss the results.

**Result:**

The maximum long number which have taken is 10^7, that to with a randomly generated key,value pairs.

The below are some randomly generated values.

Method | Data structure | Values type | Values Generated | Max Value | Time spent  
 DeDup StringSet Long 10000000 5000000 74.32  
 DeDup StringSet Long 10000000 10000000 83.83  
 DeDup StringSet Long 10000000 20000000 92.99

**Problem Statement :- 2**

Redo Exercise 3.5.30 using StringSET (see Exercise 5.2.6) instead of HashSET. Compare the running times of the two approaches. Then use Dedup to run the experiments for N = 10^7, 10^8, and 10^9, repeat the experiments for random long values and discuss the results.

**Result:**

The maximum long number which have taken is 10^7, that to with a randomly generated key,value pairs. By using dedup method and hashset data structure. The below are some randomly generated values.

Method | Values type | Values Generated | Max Value | Time spent  
  
 DeDup Long 10000000 5000000 25.69  
 DeDup Long 10000000 10000000 29.37  
 DeDup Long 10000000 20000000 33.38

The time complexities dedup method is O(N), add method in HashSet is O(1), contains method in HashSet O(1).

**Problem Statement:- 3**

Write a client that takes integers M, N, and T as command-line arguments,  
then uses the code given in the text to perform T trials of the following experiment: Generate N random int values between 0 and M – 1 and count the number of duplicates. Run your program for T = 10 and N = 10^3, 10^4, 10^5, and 10^6, with M = N/2, and N, and 2N. Probability theory says that the number of duplicates should be about (1 – e^(–alpha) ) where alpha = N/M—print a table to help you confirm that your experiments validate that formula.

**Result:**

As per the problem statement, I have used (1 – e^(–alpha) ) for distinct values, by taking the mean of all T trails. Values for N = 10^3, 10^4, 10^5, and 10^6, with M = N/2, and N, and 2N have approximately equal values.

Values Generated | Max Value | Distinct Values | Expected Distinct Values  
  
 1000000 500000 432384.60 432332.36  
 1000000 1000000 632184.60 632120.56  
 1000000 2000000 786948.20 786938.68

The time complexity for the problem statement methods is O(T\*(M+N))