For experimental analysis, briefly describe your experimentation framework (e.g., what synthetic data you have generated and why, what you have measured)

For experimental analysis on synthetic data, we construct six experiments on different aspects: number of strings, inserting order, length of each string, proportions of duplicates, number of none values, and searching values not in the set. We measure the inserting time for the first five experiments, and searching time for all experiments.

For number of strings, we construct different number of strings with length 5-10 characters to verify growth in time.

For inserting order, we insert the same data in ascending, descending or random order because this affects algorithms’ behaviors, especially for binary search tree.

For length of each string, we use strings with different lengths to check if longer strings lead to longer time due to complexity in comparison.

For proportions of duplicates, we repeated different proportions of data to construct duplication.(thoroughly shuffled) Since set does not contain duplicated keys, this would lead to a smaller size for data structures and thus reduce time taken.

For number of none values, we included different number of none values. (thoroughly shuffled) This is similar to duplicates, where none values do not need to be inserted and the size reduces.

For searching values not in the set, we construct proportions of the searching values using a different set of random strings from the inserting ones. (thoroughly shuffled) (There is a very small probability that same random strings appeared in both inserting and search values. However, this is negligible). We did this because we want to verify if searching values not in the set takes longer than searching values in the set.