Lab Assignment 6 (Perfect Hashing): CS2233

16th October, 2023

Datasets: Please use **T10I4D100K**, dataset mentioned at the link below[1]. The dataset consists of streams of integers. You may consider all the numbers mentioned in the file as one stream. Also, please consider the universe as the set of positive integers.

- Implement `perfect hashing` algorithm discussed in the class. Create a hash table corresponding to the given dataset. Consider the first 5% of elements of the dataset, and generate a set of random integers of the same size (5% of the dataset), take the union of these two sets as your test datasets, verify and report if your table is outputting the correct result. Perfect hashing consider $(key, value)$ to build the hash table. Suppose x is an element in the dataset, you can consider (x, x) as your key value pair. Report the confusion matrix on the test set.

---

[1] http://fimi.uantwerpen.be/data/