Project Proposal

Advanced Data Structure
Sadiya Ahmad
Computer Science

Van Emde Boas Tree

It is a tree data structure which implements an associative array with m-bit integer keys. It performs all operations in O(log log M) time where M is the maximum number of elements that can be stored. An extension to Van Ende Boas tree is to store max and min values in it for each cluster of the tree. Lets us first clearly explain what is a cluster?

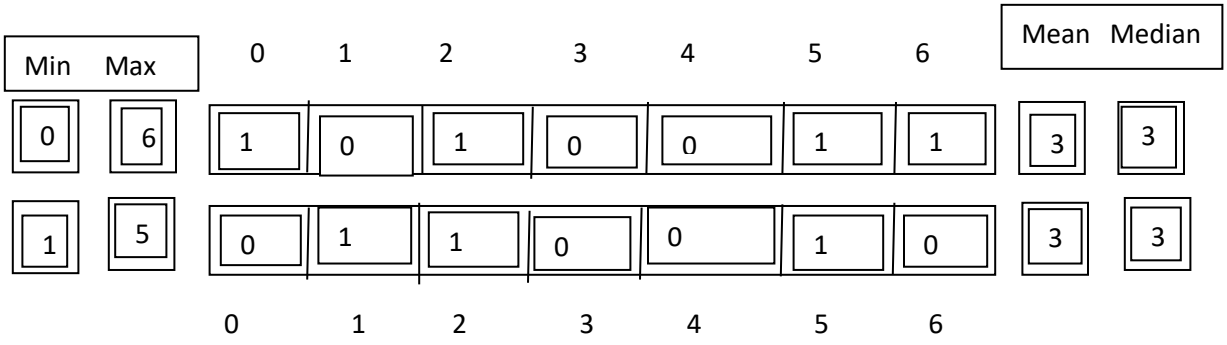
Let us assume we have a universe of elements u. We create a bit vector array of u elements and divide it into clusters(galaxies) of size sqrt(u). Suppose elements present in U are {1,9,10,15} then we will insert 1 in index 1, 1 in index 9, 1 in index 10 and 1 in index 15. All other index will be 0. Now consider converting this array into double dimension array. Each row will have sqrt(u) elements. These rows are called as clusters(galaxies). An extension to this fancy conversion is to add minimum and maximum values for each row. Now suppose we need to insert an element 8 in the data structure. We can do following for that purpose:

8 = 10 00

First two digits(10) are called as H(x) and last two digits(00) ae considered as L(x). For inserting an element we find the row of H(x) i.e row 1 and column of L(x) i.e column 0 . Now [1][0] in array is assigned 1. Max and min values are stored for each row or cluster so as to minimize the search size.

Proposed idea:

As we know we are drastically decreasing the search time by pre storing min and max, our proposal is to store mean and median for the elements in the universe u. This can be helpful to plot graphs based on the elements of u and not make any search when a graph is requested. The update in mean and median can be made as per the changes in max and min. Let us take an example where we have two clusters of size 7. Suppose S = {0,2,5,6,8,9,12}



Project Proposal

Advanced Data Structure
Sadiya Ahmad
Computer Science

Cluster 0 Mean = 0 + 3 = 00 + 11 = 0111 = 3

Max stores the index value of the last occurring element and Min stores the index value of first occurrence of the element. These mean and median can change with the change in Min and Max of each cluster.

Once the implementation of insert and delete have been performed keeping mean and median under consideration we can observe the time complexity of running this algorithm. There will test run on different data set to get the approximate time.

Further this can be extended to include standard deviation. With so many parameters at hand in the data structure it would become quite easy to plot a graph for each cluster, in fact even for all the value sin the universe. As we know Van Emde boss Tree is widely used in routers to send packets from one destination to another, the analysis of these transfer will become easier and convenient. This implementation can be carried out either in C++ or java depending on ease of the language.