**Documentation Project**

**Advanced Algorithm and Data Structures**

Group XX (non so che numero siamo)

Exercise 1:

*Import on class Node:*

* **from b\_tree.sorted\_table\_map import SortedTableMap**
* **Math**

The exercise requires the implementation of a balanced\_tree. To implement the node structure we have been developed the class **Node** that has these attributes:

* **stm:**

This attribute is an object of class **SortedTableMap.** This sorted collection stores keys and relative values.

* **\_parent:**

It contains the address of the parent’s node.

* **\_middle:**

This is the most important attribute of the class. It contains the middle element (couple key-value) of the instance of the node.

* **\_d:**

Another important attribute. It define max number of the element storage in the node, therefore also the frequency of the overflow’s/underflow’s operation.

* **overflow/underflow:**

These attributes represent the state of the node (True if equal to one, False if equal to zero).

Furthermore the class Node has these methods:

* **\_\_init\_\_(self, k, v, p, d)**

This is the constructor of the class. k-v are the couple key-value of the first element of the node, p is the address of the node parent and d is the max number of elements of the node (d is always ten for each node). It calls the \_middle\_element\_ method.

* **\_middle\_element(self):**

This method find the middle element of the node using a dichotomic search and insert it on the **\_middle** attribute. If the length of the stmtable map is one the value will be the first and only element of the map (the method use the **Math** library).

* **get\_middle\_element(self):**

This method return the middle element of the node calculate with the private method.

* **insert\_element(self, k, v):**

This method insert the couple k-v in the map and update the overflow, underflow and \_middle attributes.

* **delete\_element(self, k):**

This method delete the couple in the map where the key is equal to k and update the underflow and \_middle attributes.

* **print\_node(self):**

This method print all the elements in the map of the node and after that print the \_middle attribute.

With this structure the insert and delete operation have a complexity of O(log(\_d)) using the dichotomic search and where \_d is the value of the relative attribute.

*Import on class MultiWaySearchTree:*

* **from b\_tree.sorted\_table\_map import SortedTableMap**
* **from b\_tree.node import Node**
* **Math**

This class represent the whole structure of the tree and it has these attributes:

* **tree:**

Is a SortedTableMap where the key of a couple key-value is the middle\_element of a node and the value of couple is the relative node of the middle\_element.

* **node:**

This attribute is an object of the class Node and it represent the node that we consider at this instant.

* **root:**

This is a tuple that contain the couple key-value of the root node of the tree.

Moreover the class MultiWaySearchTree has these method:

* **\_\_init\_\_(self, k, v):**

The constructor of the class. Create a new node with couple k-v and parent’s address equal to None. After that insert the node in the map and call the \_root\_node method.

* **is\_root(self):**

This method return True if node is equal to the root.

* **\_root\_node(self):**

This method calculate what node is the root and fill the root attribute with the result couple key-value. If the length of the map is one the method will set root with the first and only element of the map.

* **is\_empty(self):**

This method return True if the length of the node attribute is zero.

* **find\_node(self, k, type):**

This method find the correct node where to insert or delete the element. The parameter type is True when this method is invoked from the new\_element method and False when is invoked from the delete\_tree\_element method. The node will be convert from tuple to list. This method has a complexity of O(log n) where n is the number of nodes of the three.

* **new\_element(self, k, v):**

This method insert the element in the correct node using the result of the find\_node method and using the insert\_element method of the class Node. After that invoke the fix\_node method and if the overflow attribute of the node is equal to one call the \_split method.

* **fix\_node(self):**

This method is very important because represent the bottle neck of the whole tree computational complexity. Whenever that new\_element and delete\_tree\_element methods are called, we iterate the map of the node and we update the key in the couple key-value of the node used by the insert/delete operation. This method have complexity of O(n) where n is the number of nodes of the tree.

* **\_split(self, node):**

This method is called from the new\_element method when there is an overflow on the node. It splits the overflow’s node in two nodes (therefore create new nodes and insert in the map) when the overflow’s node is not the root and in three nodes when is the root. Moreover this method invoke the fix\_node method to recalculate the key in the map.

* **delete\_tree\_element(self, k):**

This method delete the couple key-value in the correct node (using the search\_other\_node method). After the delete of the key, this method call the fix\_node method and the verify\_underflow method to avoid wrong key’s value in the map and underflow on the node where we deleted the element.

* **search\_other\_node(self, k, direction):**

This method is similar to the find\_node method, but if direction is True search on the right side of the three and if False search on the left side of the three. Moreover it find the first node that is not the node that contain the k value and we used this method to avoid the delete into a node that is not a leaf.

* **verify\_underflow(self, k, node):**

This method verify if there is an underflow on the node and call the \_transfer or \_fusion method using the length of the adjacent node (this method call the find\_first\_util\_node and search\_siblings methods).

* **find\_first\_util\_node(self, k, direction):**

This method find the father node of the node with key equal to k. If direction is True search the node on the right side of the three, if False search on the left side of the three. In the case that we don’t find father in the cycle, the father is the root and we return it.

* **search\_siblings(self, k, direction):**

This method find the adjacent brother of a node that have key equal to k. If direction is True search for the right brother, if False search for the left brother. Return None if we don’t find brother.

* **\_transfer(self, brother, direction, father, istance):**

This method do the exchange for the brother-father-istance. Istance generated an underflow, to fix it if the direction is True (False) the max value of father will be copy on istance and deleted from father. After that the max value of the brother will be copy on father and deleted from brother. We call also the fix\_node method. If direction is True the brother will be the left brother of istance, if False will be the right brother of istance.

* **\_fusion(self, brother, direction, father, istance):**

This method is similar to the \_transfer method, but istance will be deleted (fusion with brother) and the max value of the father will be copy on brother and deleted from father.

* **print\_tree(self):**

This method print all the couple key-value and their relative node.

The computational complexity for the insert/delete depends on the overflow/underflow operation and on the method fix\_node. The complexity will increase at the grow of the number of nodes and will decrease at the grow of the number of the elements in the single node.

Exercise 2:

*Import:*

* **time**
* **random**
* **from** **priority\_queue.adaptable\_heap\_priority\_queue import AdaptableHeapPriorityQueue**

The exercise requires the implementation of a non pre-empitive scheduler. It’s has been developed the class Scheduler that inherits from **AdaptableHeapPriorityQueue**. It also stores the number of time slices after which the priority of a task in the scheduler itself must be incremented. The scheduler class has the following public methods:

* **\_\_init\_\_(self,slice\_to\_increment):**

The constructor of the class. It requires the parameter slice\_to\_increment.

In this method the super() constructor is called, the value slice\_to\_increment is saved and it’s defined another attribute of the class: number\_time\_slice that indicates how many time slices have been spent.

* **add\_job(self,k,v):**

This method allows the insertion of a new job in the scheduler. It requires as input parameters, the priority of the task *k* and a tuple *v* that contains the name of the job and how many time slices it requires. This method verifies that the priority of the job is a valid number (otherwise an Exception is raised), re-arrange the information present in the tuple and insert the job into the Scheduler via the *add*() method present in **AdaptableHeapPriorityQueue.**

* **job\_execution(self):**

This method prints on the output information about the current job in execution, if the scheduler is empty the message “*The scheduler has no tasks.”* is printed. This method is responsible for the update of the priority of the job, the method *update*() in the class **AdaptableHeapPriorityQueue** is used.

In the same script is also present a function, **random\_add(scheduler)**, that randomly adds a job in the scheduler. The job could be chosen among a list: *CPU*, *Memory* and *I/O.* The priority of each task is randomly generated and even the number of time slices for each task is randomly generated. When the script is run the first time, it requires as input the number of time slices after which the priority must be incremented. A check is present to verify that the user has inserted a valid number. After this “setup” phase, five jobs are added to the scheduler. In the infinite loop, it is generated a random value and if it is greater than a threshold, a new task is added to the scheduler.

Exercise 3:

*Import:*

* **from TdP\_collections.graphs.my\_graph import My\_graph**

The exercise requires the implementation of an iterative DFS algorithm without using an auxiliary data structure. It’s has been developed modifying the Graph’s Vertex class. It has been added a parameter: **\_pre**. The **Vertex** class has the following public methods (there are other methods which are used to solve exercise 5 so they are not illustrated here):

* **\_\_init\_\_(self,x)**: the constructor of the class. It requires the parameter x that is the element that must be instanced.
* **element(self)**: this method allows to return the element of the **Vertex**.
* **pre(self):** this method allows to return the previous element from which the current **Vertex** is discovered.
* **set\_pre(self,x=None)**: this method allows to modify the attribute pre of the **Vertex**.

The script takes in input the graph on which the DFS must be performed and the source vertex. The discovered vertices are stored in a list, so the first element of this list is the source vertex. The algorithm, iterating on all the incident edges of the source vertex, checks if the other vertex (not the source one) has been already visited, with the **\_pre** attribute. If the **Vertex** has not been visited it is stored in a variable, the \_**pre** attribute is and it is added to the nodes list, otherwise continue. ####Now for the successive vertex is checked if any of the incident edges there is a not visited vertex, if there are the algorithm continue to go deeper, if all near vertex are visited the algorithm go back and search again until all the vertex of the vertex are visited and the next incident edge of the source node is checked. The algorithm stops when all the incident edge of the source node are checked and returns a list containing all the connected nodes of the graph.

Exercise 4:

*Import:*

* **from tree.concrete\_tree import ConcreteTree**

This exercise requires us to implement a Dynamic Programming algorithm that’s able to install a software on the minimum number of users of a social network so that for every pair of friends, at least one of them has the software. The network is modelled as a tree. From now on, the sentence “color a node” becomes equivalent to “install a software for a given user”. In order to solve this problem, we have provided our implementation of the **Tree** class, the **ConcreteTree** class. It contains the **\_Node** class with the following attributes:

* **\_element:** the element stored in the node;
* **\_parent:**  a reference to the node’s parent;
* **\_children:** a list that contains all the children of the node;
* **\_colored:** an attribute to specify if the node must be colored or not;
* **\_pos:** the position of the **node** in the array.

The **\_Node** class also contains the method:

* **\_\_init\_\_(self,element,parent=None,children=None):** it’s used to instantiate the class.

The class **ConcreteTree** also contains the class **Position** which inherits from the **Tree.Position** class. This class has the following methods:

* **\_\_init\_\_(self,container,node):** this method is the constructor for the class;
* **element(self):** this method returns the element stored in the node;
* **get\_parent\_position(self):** this method returns the position, in the list, of the node’s parent;
* **node(self):** returns the node stored in the position;
* **color(self):** color the node;
* **set\_array\_position(self,index):** this method is used to store the position of the node in the list;

**\_\_str\_\_(self):** this method is implemented in order to print, when the tree is visited, the element stored in the position and if that position is colored or not;

* **\_\_eq\_\_(self,other):** this method checks if this position and the other position are the same position.

The class **ConcreteTree** contains the following methods:

* **\_\_init\_\_(self):** this is the constructor of the class;
* **\_\_len\_\_(self):** returns the number of the elements present in the tree;
* **root(self):** returns a **Position** that represents the root of the tree, or None;
* **parent(self,p):** returns the **Position** of the node that is the parent of the node stored in p;
* **children(self,p):** returns a list of **Position** in which are stored the children of the node in the **Position** p;
* **\_add\_root(self,e):** this method allow us to insert as root the element e;
* **\_add(self,p,e):** this method adds, as child of p, a new **Node** that stores the element e;
* **num\_children(self,p):** this method returns the number of children of the **Position** p;
* **\_validate(self,p):** returns the node stored in the **Position** p, if it’s valid;
* **\_make\_position(self,node):** returns the **Position** in which the node is encapsulated.

The idea behind the algorithm is traverse the tree in a preorder way. For each element, we set the position in the list as attribute of the *i-th* **Position**, and we store, in the list, the *i-th* **Position** and the number of its children those are not colored (we later explain why). Then we iterate from the last element of the list to the first element and for each element we check the number of its children:

* number of children less or equal to zero: continue the execution;
* number of children greater than zero: we color it, we get the position of the parent of the element because we decrease the number of the non-colored children (if a node has all children colored we don’t need to color it) and go to the next iteration.

This algorithm is obtained from a similar algorithm that uses the recursion to solve the problem. The base cases for the recursive algorithm are:

* empty tree;
* tree with one node;
* tree with height equal to 1.

In the first two cases we don’t need to color any node, in the last case we need to color only the root. Starting from a node the algorithm is executed on its children; the starting point is the tree’s root.

To transform this algorithm in a DP algorithm we have introduced an auxiliary data structure and we have chosen a list in order to respect the specific about time complexity. The algorithm is explained above.

Exercise 5:

*Import:*

* **random**
* **time**
* **from graphs.my\_graph import My\_graph**

This exercise requires us to implement the solution of the previous exercise considering the BaceFook network as a graph and to implement the solution with a greedy algorithm. First, we have implemented the class **My\_graph** which inherits from the **Graph** class in the Tdp.collection package. This class adds, as attribute of the Vertex inner class (which inherits from the **Graph.Vertex** class), the element ***colored.*** This element is used to mark if a Vertex of the graph should have the software or not. The class **My\_graph.Vertex** also has the methods:

* **\_\_init\_\_(self,x,colored=False):**

this method has, as input variables, the element that should be stored and the colour for the vertex. It calls the **Graph.Vertex** constructor via the super().\_\_init\_\_() method, which has, as input, the value that is stored in the graph. The method also assigns the value of the colored parameter to the colored slot.

* **colored(self):**

this method returns the color associated with the specific vertex.

* **Color(self):**

this method colors the specific vertex.

The **class My\_graph** instead has the following methods:

* **not\_colored\_vertex(self,v):**

this method is a generator. It iterates over all the vertices connected to a vertex v (after the check that v is a valid vertex), that is the input parameter, those are not colored.

* **get\_vertices(self):**

this method returns a list of all the vertices of the graph.

* **not\_colored\_degree(self,v):**

this method computes the difference between the total number of the incident edges of a vertex and the number of incident edges that, at the opposite of v, have already a colored vertex.

* **get\_vertex(self,value):**

this method has, as input parameter, a value. It looks for the vertex which stores that specific value.

The function **color\_vertex** implements the algorithm: its input is a graph. We iterate among all the vertices in the graph, so the specific vertex could be:

* colored, in this case we skip and pass to the next vertex;
* not colored, in this case we iterate among all the vertices connected to this vertex. If the first vertex has a “*not\_colored\_degree”* higher than the second one we color the first one and start with a new vertex, otherwise we color the second vertex and continue among the remaining vertices.

The script “Exercise5.py” also contains the generation of 100 graphs of 100 vertices, vertices are connected randomly. The proposed solution is then applied on all these graphs and, each time an execution ends, we print information about the execution time in nanoseconds and the number of colored graphs. The optimum solution for this problem is quiet difficult to express because it depends also on the topology of the graph (in a “star” graph if we color just one node we have achieved the optimum) but we can make some theoretical consideration: this solution, because is a greedy solution, is 2-approximated (according to the Graham’s Theorem) as upper bound and, because our problem is analogous to a weighted vertex cover problem we know that, according to the Dinur-Safra’s Theorem, there is, as upper bound, a 1.3-approximation solution, so we can certainly say that the performance of our algorithm is between these two bounds.