

Μυλωνάκης Αλέξανδρος

AM: 3045

Εργασία 2^η Διαχείριση Σύνθετων δεδομένων

Γλώσσα προγραμματισμού : Python

Μέρος 1^ο:

Κατέβασα το αρχείο tiger.csv και δημιούργησα την κλάση csv_manager.py για να διατρέχω το αρχείο και να αρχικοποιώ την λίστα της οποίας το κάθε αρχείο περιέχει στην 1^η θέση το identifier του αντικειμένου, στην 2^η τα min_max του αρχείου και στην 3^η το linestring

```
def create_mylist(self):  
    road_list = [[] for i in range(3)]  
  
    with open(self.csv_path) as csvfile:  
        reader = csv.reader(csvfile)  
        identifier = 1  
        for row in reader:  
            if len(row) > 1:  
                road_list[0].append(identifier)  
                road_list[1].append(self.find_line_minmax(row))  
                road_list[2].append(row)  
                identifier+=1  
    return road_list
```

Η κλάση αυτή επίσης είναι υπεύθυνη να για να βρίσκει τα ελάχιστα και μέγιστα X και Y από όλα τα linestrings έτσι ώστε να φτιάξουμε το grid

```

def __init__(self, csv_path):
    self.csv_path= csv_path
    self.minX =float('inf')
    self.minY=float('inf')
    self.maxX=float('-inf')
    self.maxY=float('-inf')

def find_line_minmax(self,row):
    minX =float('inf')
    minY=float('inf')
    maxX=float('-inf')
    maxY=float('-inf')
    for column in row:
        values = column.split()
        x = float(values[0])
        y = float(values[1])
        if(minX > x):
            minX=x
            if(x<self.minX):
                self.minX=x
        if(minY > y):
            minY=y
            if(y<self.minY):
                self.minY=y
        if(maxX < x):
            maxX=x
            if(x>self.maxX):
                self.maxX=x
        if(maxY < y):
            maxY=y
            if(y>self.maxY):
                self.maxY=y
    s1 = '{} {}'.format(minX, minY)
    s2 = '{} {}'.format(maxX, maxY)
    s3 = [s1,s2]
    return s3

```

Αυτό το πετυχαίνουμε αρχικοποιώντας τόσο τα τοπικά όσο και τα ολικά μέγιστα και ελάχιστα σε τιμές -άπειρο και +άπειρο αντίστοιχα.

Στη συνέχεια για να δημιουργήσω τη σχάρα έφτιαξα την κλάση `my_grid.py` η στην οποία αρχικοποιώ τα `minmax` στοιχεία το υψος και το παχος των κελιών

```
def __init__(self, road_list, csv_manager, number_of_rows):
    self.road_list = road_list
    self.csv_manager = csv_manager
    self.minX = csv_manager.getMinX()
    self.minY = csv_manager.getMinY()
    self.maxX = csv_manager.getMaxX()
    self.maxY = csv_manager.getMaxY()
    self.cell_width = (self.maxX - self.minX) / number_of_rows
    self.cell_height = (self.maxY - self.minY) / number_of_rows
    self.grid_map = {}
    self.number_of_rows = number_of_rows
    for y_coord in range(number_of_rows):
        for x_coord in range(number_of_rows):
            coord = '{},{}'.format(y_coord, x_coord)
            self.grid_map[coord] = []
```

Και τη μέθοδο `create grid` για την δημιουργία του grid και των αρχείων `grid.dir` και `grid.grd`

```
def create_grid(self):
    for row in range(len(self.road_list[0])):
        self.find_MBR_cells(self.split_minXY_maxXY(self.road_list[1][row]), self.road_list[0][row])

    with open('C:\\Users\\Alekos\\Desktop\\projects\\Diaxeirisi\\assignment2\\grid.grd', 'w') as f:
        for key in self.grid_map:
            value = self.grid_map[key]
            if len(value) > 0:
                for element in range(len(value)):
                    x = value[element]
                    identifier = self.road_list[0][x-1]
                    mbr_str = ' '.join(map(str, self.road_list[1][x-1]))
                    vertex_str = ','.join([''.join(map(str, vertex)) for vertex in self.road_list[2][x-1]])
                    f.write('{},{},{}\n'.format(identifier, mbr_str, vertex_str))

    with open('C:\\Users\\Alekos\\Desktop\\projects\\Diaxeirisi\\assignment2\\grid.dir', 'w') as f:
        f.write('{} {} {} {} '.format(self.minX, self.maxX, self.minY, self.maxY))
        for key in self.grid_map:
            elements = len(self.grid_map[key])
            x_coord = (key[1])
            y_coord = (key[3])
            f.write('\n{} {} {} '.format(x_coord, y_coord, elements))
```

Η συνάρτηση find_MBR_cells είναι υπεύθυνη για να βρίσκει τα όρια του MBR του

```
def find_MBR_cells(self, elem_minXY_maxXY, index):
    bottom_left = self.find_vertice_cell( elem_minXY_maxXY[0], elem_minXY_maxXY[1])
    top_right = self.find_vertice_cell( elem_minXY_maxXY[2], elem_minXY_maxXY[3])
    for row in range(bottom_left[1],top_right[1]+1):
        for col in range(bottom_left[0],top_right[0]+1):
            map_key = '({},{})'.format(row,col)
            self.grid_map[map_key].append(index)
```

Επίσης έχω υλοποιήσει την συνάρτηση fid_vertice_cell που είναι υπεύθυνη για να βρίσκει το κελί που βρίσκεται μια κορυφή

```
def find_vertice_cell(self,x,y):
    row = int((y - self.minY) // self.cell_height)
    col = int((x - self.minX) // self.cell_width)
    if(col == self.number_of_rows):
        col -=1
    if(row == self.number_of_rows):
        row -=1
    cell_value = [row,col]
    return cell_value
```

Μέρος 2^ο

Όσο αναφορά το 2^ο μέρος δημιούργησα μια κλάση `manage_files.py` για να διαβάσω τα αρχεία `grid.dir`

```
def read_file_and_split(self,file_name):  
    with open(file_name, 'r') as f:  
        contents = f.read()  
        lines = contents.splitlines()  
        return lines
```

και `grid.grd`

```
def read_grd_file(self,file_name):  
    with open(file_name) as f:  
        my_list = []  
        for line in f:  
            fields = line.strip().split(",")  
            my_grid_to_list = []  
            my_grid_to_list.append(fields[0])  
            my_grid_to_list.append(fields[1])  
            my_grid_to_list.append(fields[2:])  
            my_list.append(my_grid_to_list)  
    return my_list
```

Η κλάση αυτή επίσης περιέχει δυο ακόμη μεθόδους την `read_queries` για να διαβάζει το αρχείο `queries.txt`

```
def read_queries(self,file_name):  
    with open(file_name, 'r') as f:  
        contents = f.read()  
        lines = contents.splitlines()  
        for i in range(len(lines)):  
            lines[i] = lines[i][2:]  
    return lines
```

Και μία μέθοδο για να κάνει split lines με 4 στοιχεία

```
def split_line(self,line):  
    n1,n2,n3,n4 =line.split()  
    n1 = float(n1)  
    n2 = float(n2)  
    n3 = float(n3)  
    n4= float(n4)  
    return n1,n2,n3,n4
```

Στη συνέχεια δημιούργησα την κλάση selection_queries.py στην οποία αρχικοποιώ πάλι τα min_max to grid map μου καθώς και την συνάρτηση find_vertice_cell

```
def __init__(self,number_of_rows):  
    self.minX =float('inf')  
    self.minY=float('inf')  
    self.maxX=float('-inf')  
    self.maxY=float('-inf')  
    self.cell_width = 0  
    self.cell_height = 0  
    self.file_manager = manage_files()  
    self.grid_map = {}  
    self.number_of_rows = number_of_rows  
    for row in range(number_of_rows):  
        for col in range(number_of_rows):  
            coord = '({},{})'.format(row,col)  
            self.grid_map[coord] = []  
  
    def find_vertice_cell(self,x,y):  
        row = int((y - self.minY) // self.cell_height)  
        col = int((x - self.minX) // self.cell_width)  
        if(col == self.number_of_rows):  
            col -=1  
        if(row == self.number_of_rows):  
            row -=1  
        cell_value = [col,row]  
        return cell_value
```

Έφτιαξα μια συνάρτηση `create_structure` για τη δημιουργία της δομής από την επεξεργασία των αρχείων `grid.dir` και `grid.grd`

Η δομή που χρησιμοποίησα όπως και στην δημιουργία του `grid` στο πρώτο μέρος (ένα dictionary)

```
def create_structure(self, my_grid, my_grid_dir):
    flag = False
    counter=0
    for line in my_grid_dir:
        if not flag:
            flag = True
            self.minX, self.maxX, self.minY, self.maxY = self.file_manager.split_line(line)
            self.cell_width = (self.maxX - self.minX) / self.number_of_rows
            self.cell_height = (self.maxY - self.minY) / self.number_of_rows
            continue
        y,x,elements = line.split()
        key = '({},{})'.format(y,x)
        if int(elements)==0:
            continue
        for i in range(int(elements)):
            self.grid_map[key].append(my_grid[counter])
            counter += 1
```

Μέρος 2^ο και 3^ο

Έφτιαξα την μέθοδο `selection_query` για την αρχικοποίηση του παραθύρου(από το `query`) και για την επεξεργασία τους

```
def selection_query(self, wx_min , wx_max, wy_min, wy_max):
    bottom_left = self.find_vertice_cell(wx_min, wy_min)
    top_right = self.find_vertice_cell(wx_max, wy_max)
    identifiers = []
    refined_identifiers = []
    cells = 0
    for row in range(bottom_left[0],top_right[0]+1):
        for col in range(bottom_left[1],top_right[1]+1):
            coord = '{},{}'.format(row, col)
            elements = self.grid_map.get(coord)
            if len(elements)==0:
                continue
            cells += 1
            for element in elements:
                mbrminX,mbrminY,mbrmaxX,mbrmaxY = self.file_manager.split_line(element[1])
                if not self.check_window_intersection(wx_min , wy_min, wx_max, wy_max,mbrminX,mbrminY,mbrmaxX,mbrmaxY):
                    continue
                ref_x,ref_y =self.find_refernce_point(mbrminX,mbrminY,mbrmaxX,mbrmaxY)
                if self.point_in_window(ref_x,ref_y,row,col):
                    if self.refinement_step(wx_min , wx_max, wy_min, wy_max,element[1],element[2]):
                        refined_identifiers.append(element[0])
                    identifiers.append(element[0])
    return identifiers,refined_identifiers,cells
```

Για να ελέγξω αν επικαλύπτεται το παράθυρο από κάποιο

MBR αρχικά δημιουργώ την μέθοδο
`check_window_intersection`

```
def check_window_intersection(self,wx_min , wy_min, wx_max, wy_max, mbrx_min ,mbry_min , mbrx_max , mbry_max):
    if wx_max < mbrx_min or wx_min > mbrx_max:
        return False
    if wy_max < mbry_min or wy_min > mbry_max:
        return False
    return True
```

Για να βρω το `reference_point` έφτιαξα την

`find_reference_point`


```

def find_refernce_point(self,mbrX,mbrY>windowX>windowY):
    if mbrX>windowX:
        x = mbrX
    else:
        x = windowX
    if mbrY>windowY:
        y= mbrY
    else:
        y=windowY
    return x,y

```

Για να δω αν κάποιο σημείο βρίσκεται μέσα στο παράθυρο

```

def point_in_window(self,ref_x,ref_y,row,col):
    cell_value = self.find_vertice_cell(ref_x,ref_y)
    if cell_value[0]==row:
        if cell_value[1]==col:
            return True
    return False

```

Τέλος για το refinement_stage έφτιαξα την refinement_step η οποία επιστρέφει μια Boolean τιμή αν στο παράθυρο περιέχεται ένα linestring

```

def refinement_step(self,wx_min , wx_max, wy_min, wy_max,mbrminmax,linestring):
    mbrminX,mbrminY,mbrmaxX,mbrmaxY = self.file_manager.split_line(mbrminmax)
    if wx_min< mbrminX and wx_max > mbrmaxX:
        return True
    if wy_min < mbrminY and wy_max > mbrmaxY:
        return True
    if self.line_points_interesction(wx_min , wx_max, wy_min, wy_max,linestring):
        return True
    return False

```

Αυτό το πετυχαίνει με την βοήθεια των μεθόδων `line_points_intersection` και `check_intersection` για την περίπτωση που δεν είναι επικαλυπτόμενα σε κάποιον άξονα τα παράθυρα εκ των οποίων η `check_intersection` ελέγχει αν υπάρχει σημείο τομής μεταξύ μιας πλευράς και του `linestring` και η `line_points_intersection` διατρέχει το `linestring`.

```
def check_intersection(self,x1,y1,x2,y2,point3,point4):
    x3,y3 = point3.split()
    x4,y4 = point4.split()
    x3 = float(x3)
    y3 = float(y3)
    x4 = float(x4)
    y4 = float(y4)
    t=-1
    u=-1
    if (x1-x2)*(y3-y4)-(y1-y2)*(x3-x4)!=0:
        t = ((x1-x3)*(y3-y4)-(y1-y3)*(x3-x4))/((x1-x2)*(y3-y4)-(y1-y2)*(x3-x4))
    if (x1-x2)*(y3-y4)-(y1-y2)*(x3-x4) !=0:
        u = ((x1-x3)*(y1-y2)-(y1-y3)*(x1-x2))/((x1-x2)*(y3-y4)-(y1-y2)*(x3-x4))
    if t>=0 and t<=1 and u>=0 and u<=1:
        return True
    return False

def line_points_interesction(self,wx_min , wx_max, wy_min, wy_max,linestring):
    prev_point =linestring[0]
    for point in linestring:
        if point == linestring[0]:
            continue
        if self.check_intersection(wx_min,wy_min,wx_min,wy_max,prev_point,point):
            return True
        if self.check_intersection(wx_min,wy_min,wx_max,wy_min,prev_point,point):
            return True
        if self.check_intersection(wx_max,wy_max,wx_min,wy_max,prev_point,point):
            return True
        if self.check_intersection(wx_max,wy_max,wx_max,wy_min,prev_point,point):
            return True
        prev_point = point
    return False
```

Testing:

Για τον έλεγχο των αποτελεσμάτων αρχικά χρησιμοποίησα το αρχείο queries.txt και τα αποτελέσματα

Pre-refinement

```
Query 1 results:
13151 15262 15774 16782 21379 22260 22500 22946 22947
Cells: 1
Results: 9
-----
Query 1 results:
33887 34512 34862
Cells: 1
Results: 3
-----
Query 1 results:
30397
Cells: 2
Results: 1
-----
Query 1 results:
1108
Cells: 1
Results: 1
-----
Query 1 results:
5496
Cells: 3
Results: 1
-----
```

Post-refinement:

```
Query 1 results:
21379 22260 22500 22946 22947
Cells: 1
Results: 5
-----
Query 2 results:
34512
Cells: 1
Results: 1
-----
Query 3 results:
30397
Cells: 2
Results: 1
-----
Query 4 results:

Cells: 1
Results: 0
-----
Query 5 results:
5496
Cells: 3
Results: 1
-----
```

Και κάποια δικά μου τεστάκια:

Στα queries

```
6, -87 -86.9 32.9 33
7,-87.752641 -87.752641 31.963678 31.963678
8, -86 -86 32 32
```

```
Query 6 results:
2800 2801 2802 2824 2825 2826 2869 2881 2904 2905 2907 2918 2919 2920 2921 2922 2923 2924 2925 2926 2927 2928 2929 2930 2931 2932 2933 2957 2972 3116 3147 3148 3217 3218 3239 3240 3241 3242 3243 3244 3245 32
46 3247 3248 3249 3250 3251 3252 3253 3255 3256 3257 3291 3295 3428 3452 3458 3459 3503 3504 3505 3515 3516 3521 3524 3525 3567 3568 3569 3570 3571 3572 3573 3574 3575 3576 3577 3578 3579 3580 3581 3582 3583
3584 3585 3586 3589 3600 3761 3784 3785 3786 3787 3848 3849 3850 3851 3870 3871 3883 3884 3885 3886 3887 3888 3889 3890 3891 3892 3893 3894 3895 3896 3897 3898 3900 3901 3904 3905 3916 3917 3929 3931 4092 4
095 4138 4141 4162 4189 4190 4210 4211 4212 4213 4214 4215 4216 4217 4219 4220 4222 4249 4254 4260 4269 4466 4467 4482 4483 4484 4501 4528 4529 4541 4542 4543 4547 4548 4549 4550 4551 4552 4553 4554 4558 476
8 4807 4848 4849 4897 4901 4903 4910 4910 4921 4922 4923 4924 4925 4926 4927 4931 4933 4934 4935 4936 4937 4938 4940 4941 4942 4946 4963 5123 5155 5156 5182 5183 5185 5195 5208 5220 5233 5234 5235 5236 5237
5238 5239 5240 5241 5242 5243 5245 5246 5251 5268 5269
Cells: 1
Results: 218
-----
Query 7 results:

Cells: 0
Results: 0
-----
Query 8 results:

Cells: 1
Results: 0
-----
```

6, -87.752641 -85.565306 31.963678 33.517448

Cells: 85
Results: 35668

Τέλος για να τρέξω τον κώδικα μου έχω υλοποιήσει μια main.py

```
manager1 = csv_manager("C:\\Users\\Alekos\\Desktop\\projects\\Diaxeirisi\\assignment2\\tiger_roads.csv")
road_list = manager1.create_mylist()
grid1 = my_grid(road_list,manager1,10)
this_grid = grid1.create_grid()

my_selection_queries = selection_queries(grid1.get_number_of_rows())
my_manager = manage_files()
my_grid_dir = my_manager.read_file_and_split('grid.dir')
the_grid = my_manager.read_grd_file('grid.grd')
my_selection_queries.create_structure(the_grid,my_grid_dir)

queries=my_manager.read_queries('queries.txt')
counter = 1

for query in queries:
    minX,maxX,minY,maxY = my_manager.split_line(query)
    identifiers,refined_identifiers,cells = my_selection_queries.selection_query(minX,maxX,minY,maxY)
    """
    identifiers_str = ' '.join(map(str,identifiers))
    #print("Pre refinement stage results")
    print("Query {} results:".format(counter))
    print(identifiers_str)
    print("Cells:",cells)
    print("Results:",len(identifiers))
    print('-----')
    #print("Post refinement stage results")
    """
    ref_identifiers_str = ' '.join(map(str, refined_identifiers))
    print("Query {} results:".format(counter))
    print(ref_identifiers_str)
    print("Cells:",cells)
    print("Results:",len(refined_identifiers))
    print('-----')
    counter += 1
```

Στην οποία τα print τα οποία βρίσκονται μέσα σε σχόλια αφορούν τον κώδικα πριν το refinement step και τα υπόλοιπα είναι οι κλήσεις των μεθόδων. Καθώς μου ζητήθηκε στην προηγούμενη άσκηση να μην δίνετε δυναμικά το path του αρχείου κράτησα την manual υλοποίηση

```
manager1 = csv_manager("C:\\Users\\Alekos\\Desktop\\projects\\Diaxeirisi\\assignment2\\tiger_roads.csv")
```

Στην οποία πρέπει να δοθεί το πλήρες path του csv αρχείου από τον προγραμματιστή στο δικό του μηχάνημα.