**МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ**

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**

**«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

Навчально-науковий комплекс

«Інститут прикладного системного аналізу»

**Кафедра системного проектування**

Лабораторна робота № 4

з курсу «Інтелектуальний аналіз даних»

**«Програмна розробка методу кластеризації даних»**

Виконала :

студентка групи ДА-42

Балан Катерина

Київ 2017

Мета роботи: Здобути навички програмної розробки методу кластеризації даних.

Завдання: Розробити програму для ієрархічної кластеризації даних.

Код програми:

*import* math  
*import* Tree  
  
train\_document = open('1.txt', 'r')  
  
train\_data = []  
*for* i *in* train\_document:  
 train\_data.append(i.lower())  
print(train\_data)  
dictionary = {}  
*for* i *in* train\_data:  
 *for* word *in* i.split():  
 dictionary[word] = 1 *if* word *not in* dictionary *else* dictionary[word]+1  
# print(dictionary)  
  
*for* key *in* list(dictionary.keys()):  
 *if* dictionary[key] == 1:  
 *del* dictionary[key]  
# print(dictionary)  
  
atribute\_y = []  
*for* i *in* train\_data:  
 count = 0  
 *for* letter *in* i:  
 *if* letter *is not* ' ':  
 count += 1  
 atribute\_y.append(count - 1)  
# print(atribute\_y)  
  
atribute\_x = []  
*for* i *in* train\_data:  
 count = 0  
 *for* word *in* i.split():  
 count += 1  
 atribute\_x.append(count)  
# print(atribute\_x)  
  
new\_data = []  
num\_sen = 0  
*for* i *in* train\_data:  
 *for* word *in* i.split():  
 *if* word *in* dictionary:  
 # print(i, word)  
 new\_data.append([word, atribute\_x[num\_sen], atribute\_y[num\_sen]])  
 num\_sen += 1  
# print(new\_data)  
  
last\_data = []  
*for* key, item *in* dictionary.items():  
 temp\_x = 0  
 temp\_y = 0  
 # print(key, item)  
 *for* i *in* new\_data:  
 *if* i[0] == key:  
 temp\_x += i[1]  
 temp\_y += i[2]  
 # print(temp\_x/item, temp\_y/item)  
 last\_data.append([key, temp\_x / item, temp\_y / item])  
print(last\_data)  
  
matrix = [[0] \* len(dictionary) *for* i *in* range(len(dictionary))]  
print(" " \* 10, end=" ")  
*for* i *in* last\_data:  
 print("%-10s" % i[0], end=" ")  
print("")  
*for* i *in* range(len(last\_data)):  
 print("%-10s" % last\_data[i][0], end=" ")  
 *for* j *in* range(len(last\_data)):  
 matrix[i][j] = round(math.sqrt(math.pow(last\_data[i][1]-last\_data[j][1], 2) +\  
 math.pow(last\_data[i][2]-last\_data[j][2], 2)), 5)  
 print("%-10f" % matrix[i][j], end=" ")  
 print(" ")

Даний код –початкова версія, яка буде допрацьована для сдачі