**МИНОБРНАУКИ РОССИИ**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
ВЫСШЕГО ОБРАЗОВАНИЯ  
**«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ  
ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ им. В.Г.ШУХОВА»  
(БГТУ им. В.Г.Шухова)**

Кафедра программного обеспечения вычислительной техники и автоматизированных систем

Лабораторная работа №11

Дисциплина: Объектно-ориентированное программирование

по теме Знакомство с Python. Основные структуры данных.

Выполнил: ст. группы ВТ-22  
Макаров Даниил Сергеевич

Проверил: Буханов Д.Г.

**Белгород 2019**

***Вариант 5***

Разработать подсистему поиска совпадений между текстами. Первая подпрограмм будет выполнять разбор и нормализацию текста. Вторая выполнять сравнение в локальном хранилище.

import re

import numpy as np

def prepare\_text(text):

#разделители для продложений

split\_sentence = re.compile(r'[.|!|?|…]')

#удаление предлогов

remove\_preposition = re.compile(r'и|или|без|под)

text = text.lower()

for char in ',-\_:;':

text.replace(char,' ')

text = re.sub(remove\_preposition,'',text.rstrip())

splited\_text = re.split(split\_sentence,text)

return splited\_text

def compare\_text(splited\_sample, splited\_text):

total\_compare\_count = len(splited\_sample)+len(splited\_text)

compare\_count = 0

for text\_str in splited\_text:

text\_list = text\_str.split(' ')

for sample\_str in splited\_sample:

sample\_list = sample\_str.split(' ')

dict = text\_list+sample\_list

text\_vector = []

sample\_vector = []

for i in range(len(dict)):

if dict[i] in text\_list:

text\_vector.append(1)

else:

text\_vector.append(0)

for i in range(len(dict)):

if dict[i] in sample\_list:

sample\_vector.append(1)

else:

sample\_vector.append(0)

result\_vector=np.array(text\_vector)+np.array(sample\_vector)

result\_sum=result\_vector.sum()

if result\_sum>len(dict)+1:

compare\_count=compare\_count+1

percent\_of\_similariry = compare\_count/total\_compare\_count\*100

return percent\_of\_similariry

print('Введите имя файла для проверки - ')

text\_file\_name = input()

print('Введите имя файла выборки - ')

ctrl\_file\_name = input()

text\_file = open(text\_file\_name)

ctrl\_file = open(ctrl\_file\_name)

text = text\_file.read()

sample = ctrl\_file.read()

splited\_text = prepare\_text(text)

splited\_sample = prepare\_text(sample)

percent = compare\_text(splited\_sample,splited\_text)

print(percent,'%')