**Национальный исследовательский университет информационных технологий, механики и оптики**

**Факультет Программной Инженерии и Компьютерной Техники**

**Лабораторная работа №4**

**По дисциплине:**

**«Информатика»**

Работу выполнила:

Студентка группы P3112

Сенина Мария Михайловна

Преподаватель:

Малышева Татьяна Алексеевна

Санкт-Петербург

2020

**Задание 1**

Мой вариант - 3

**Текст задания**

1) Реализуйте программный продукт на языке Python, используя регулярные выражения по варианту, представленному в таблице.

2) Для своей программы придумайте минимум 5 тестов. Каждый тест является отдельной сущностью, передаваемой регулярному выражению для обработки. Для каждого теста необходимо самостоятельно (без использования регулярных выражений) найти правильный ответ. После чего сравнить ответ, выданный программой, и полученный самостоятельно.

Пример тестов приведён в таблице.

Дан текст. Требуется найти в тексте все фамилии, отсортировав их по алфавиту.

Фамилией для простоты будем считать слово с заглавной буквой, после которого идут

инициалы.

Пример:

|  |  |
| --- | --- |
| Ввод | Вывод |
| Студент Вася вспомнил, что на своей лекции Балакшин П.В. упоминал про старшекурсников, которые будут ему помогать: Анищенко А.А. и Машина Е.А. | Анищенко  Балакшин  Машина |

**Мои тесты**

|  |  |
| --- | --- |
| Текст | Верный ответ |
| Маркиз Караба. С. победил в состязании на самую глупую фамилию. |  |
| Игорев З.К. пытаясь найти тему для аннотации сильно страдал. Лучшая статья по теме была,  его собственная статья подходила только для журнала "ООО Марков С.М.". | Игорев Марков" |
| Пупкин Е.М. не смог сделать лабу по информатике, потому что его регулярное выражение не могло отличить  Пупкин Е.М. от Пупкин Е.К.. | Пупкин Пупкин Пупкин |
| Мяу о.й. - сказала сестра Сенина В.М., но наше величество Голубева Х.З. ничего не поняли. | Голубева Сенина |
| Совельев А.К. Писарева Е.М. Могильный Е.К.. | Могильный Писарев Совельев |

**Код программы**

import re  
  
  
def list\_to\_string(s):  
 str1 = **" "** return str1.join(s)  
  
  
def print\_results(match, answer):  
 print(**"---------------------------------"**)  
 print(**"Наш ответ: "**, end=**''**)  
 for i in range(0, len(match)):  
 print(match[i] + **" "**, end=**''**)  
 print(**"**\n**Правильный ответ: "** + answer)  
  
  
def find\_surnames(text):  
 pattern = re.compile(**"^[А-Я]\S+"**)  
 name = re.match(pattern, text)  
 return name.group(0)  
  
  
def find\_names(text, answer):  
 pattern = re.compile(**"[А-Я]\S\*\s+[A-Я]\.[А-Я]\."**)  
 match = re.findall(pattern, text)  
 line = []  
 for i in range(0, len(match)):  
 line.append(find\_surnames(match[i]))  
 print\_results(sorted(line), answer)  
  
  
arr = []  
arr.append([**"Маркиз Караба. С. победил в состязании на самую глупую фамилию."**, **" "**])  
arr.append([**"Игорев З.К. пытаясь найти тему для аннотации сильно страдал. Лучшая статья по теме была, "  
 "а его собственная статья подходила только для журнала** \"**ООО Марков С.М.**\"**."**, **"Игорев Марков"**])  
arr.append([**"Пупкин Е.М. не смог сделать лабу по информатике, потому что его регулярное выражение не могло отличить "  
 "Пупкин Е.М. от Пупкин Е.К.."**, **"Пупкин Пупкин Пупкин"**])  
arr.append([**"Мяу о.й. - сказала сестра Сенина В.М., но наше величество Голубева Х.З. ничего не поняли."**, **"Голубева "  
 "Сенина"**])  
arr.append([**"Совельев А.К. Писарева Е.М. Могильный Е.К.."**, **"Могильный Писарев Совельев"**])  
  
for i in range(0, len(arr)):  
 find\_names(arr[i][0], arr[i][1])

Ссылка на файл в репозитории: [https://github.com/senina-m/ITMO/blob/master/Informatics\_4/lab\_4-1.py](https://github.com/senina-m/ITMO/blob/master/Informatics_4/lab_4-2.py)

**Вывод программы:**

---------------------------------

Наш ответ:

Правильный ответ:

---------------------------------

Наш ответ: Игорев Марков

Правильный ответ: Игорев Марков

---------------------------------

Наш ответ: Пупкин Пупкин Пупкин

Правильный ответ: Пупкин Пупкин Пупкин

---------------------------------

Наш ответ: Голубева Сенина

Правильный ответ: Голубева Сенина

---------------------------------

Наш ответ: Могильный Писарева Совельев

Правильный ответ: Могильный Писарев Совельев

**Дополнительное задание 1**

**Текст задания**

1) Для выполнения задания скачайте текстовый документ по ссылке, соответствующей вашему варианту. (Мой вариант – 27)

|  |  |  |
| --- | --- | --- |
| 20 … 29 | https://drive.google.com/file/d/1YoQjz0dXSz9XFrPIIBHM\_CHmJKHpYlNo/view | Ромео и Джульетта |

2) Реализуйте программный продукт на языке Python, используя регулярные выражения, выделите все предложения из вашего текста, соответствующие критерию по вашему варианту.

(Мой вариант – 3)

|  |  |
| --- | --- |
| 3 | Все вопросительные предложения, в которых слов меньше 6 |

3) Придумайте способ, как проверить корректность работы вашего регулярного выражения и проверьте его.

**Код программы**

import re  
  
  
def find\_re(text):  
 pattern = re.compile(**r"[A-Z]\w+(?:\s+\w+){0,5}\?"**)  
 return re.findall(pattern, text)  
  
  
filename\_for\_parsing = **"RomeoAndJuliet.txt"**filename\_for\_result = **"lab\_4-2\_output.txt"**file\_read = open(filename\_for\_parsing, **"r"**, encoding=**'utf-8'**).read()  
file\_write = open(filename\_for\_result, **"w"**, encoding=**'utf-8'**)  
  
line = **"**\n**"**.join(find\_re(file\_read))  
file\_write.write(line)

Ссылка на файл в репозитории: <https://github.com/senina-m/ITMO/blob/master/Informatics_4/lab_4-2.py>

**Вывод программы в файл lab\_4-2\_output.txt**

Ссылка на файл в репозитории: <https://github.com/senina-m/ITMO/blob/master/Informatics_4/lab_4-2_output.txt>

**Проверка правильности работы**

Можно либо глазами посмотреть, либо, например проверить кодом автоматически, что в каждом приложении <6 слов и в конце вопрос.

Например, таким кодом:

for i in find\_re(file\_read):  
 sentanse = i.split()  
 length = len(sentanse)  
 if length > 6:  
 print(**"Error: more than 6 words"**)  
 if sentanse[length - 1][-1] != **"?"**:  
 print(**"Error: isn't question"**)

**Дополнительное задание 2**

**Текст задания**

1) Реализуйте программный продукт на языке Python, используя регулярные выражения по варианту, представленному в таблице.

(Мой вариант - 3)

С помощью регулярного выражения найти числа, в которых в средине числа стоит 0 и затем перенести его в конец числа.

Пример

|  |  |
| --- | --- |
| Ввод | Вывод |
| 101 + 2430123 = 2430224 | 110 + 2431230 = 2432240 |

2) Для своей программы придумайте минимум 5 тестов.

3) Протестируйте свою программу на этих тестах.

**Мои тесты**

|  |  |
| --- | --- |
| Моё значение | Верное значение |
| 108791237 pkh | 187912370 |
| 190982, 987173 | 199820 |
| 00000, 0001 | 00000 0100 |
| 987028935 | 987289350 |
| kjf0jkha |  |

**Код программы**

import re  
  
  
def find\_numbers(text):  
 pattern = re.compile(**r"(?<=\s)\d+0\d(?=\s)"**)  
 return re.findall(pattern, text)  
  
  
def print\_answer(result, answer):  
 print(**"Мой ответ:"** + answer)  
 print(**"Верный ответ:"** + result)  
  
  
arr = [[**"108791237 pkh"**, **"187912370"**], [**"190982, 987173"**, **"199820"**], [**"00000, 0001"**, **"00000 0100"**],  
 [**"987028935"**, **"987289350"**], [**"kjf0jkha"**, **""**]]  
  
for i in range(0, len(arr)):  
 line\_arr =[]  
 a = find\_numbers(arr[i][0])  
 for j in a:  
 line = a  
 count = 0  
 new\_num = **""** for k in range(0, ):  
 if line[k] == 0:  
 count = count + 1  
 else:  
 new\_num = line[k] + new\_num  
 line\_arr.append(new\_num)  
 print\_answer(**" "**.join(line\_arr), arr[i][1])

Ссылка на файл в репозитории <https://github.com/senina-m/ITMO/blob/master/Informatics_4/lab_4-3.py>

**Вывод программы**

Мой ответ: 187912370

Правильный ответ:187912370

Мой ответ: 199820

Правильный ответ: 199820

Мой ответ: 199820

Правильный ответ: 987289350

Мой ответ:

Правильный ответ:

**Выводы по работе**

В ходе лабораторной работы я познакомилась с регулярными выражениями и библиотекой re для языка Phyton для их использовния.