

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО

Факультет программной инженерии и компьютерной техники

Информатика

Лабораторная работа № 3 "Регулярные выражения"

Выполнил студент:

Маликов Глеб Игоревич

Группа № Р3124

Преподаватель: Болдырева Елена Александровна

г. Санкт-Петербург

2022

Оглавление

Смайлики	3
Тесты	3
Исходный Код.....	4
Результат	4
Доп. задание №1	7
Тесты	7
Исходный Код.....	7
Результат	8
Доп. задание №2.....	11
Тесты	11
Исходный Код.....	12
Результат	13
Вывод.....	17
Список Литературы.....	18

Смайлики

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.

3) Программа должна считать количество смайликов определённого вида (вид смайлика описан в таблице вариантов) в предложенном тексте. Все смайлики имеют такую структуру:

[глаза][нос][рот].

Вариантом являются различные наборы глаз, носов и ртов.

Тесты

Файл 'sm1':

```
=</**/{P" aaaaaaaaaaaaaa  
aa  
=<{P  
=<{P" cvx ={"  
={P"aaaaaaa<aaa =<{P""=<{P"=<{P  
должно быть 5 смайликов
```

Файл 'sm2':

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Sed felis eget velit aliquet. In ornare quam viverra orci sagittis eu volutpat odio facilisis. Lacus viverra vitae congue eu. Risus at ultrices mi tempus imperdiet nulla malesuada. Condimentum id venenatis a condimentum. Tellus in metus vulputate eu scelerisque felis imperdiet. Massa vitae tortor condimentum lacinia quis vel eros donec. Aliquam eleifend mi in nulla posuere sollicitudin. Faucibus interdum posuere lorem ipsum dolor. Cursus eget nunc scelerisque viverra mauris in aliquam.

должно быть 0 смайлика

Файл 'sm3':

```
'=<{P'""=<{P" =<{P  
  
=  
<  
{
```

P

=

<{

P

= <{ P

должно быть 3 смайлика

Файл 'sm4':

должно быть 0 смайлика

Файл 'sm5':

должно быть 0 смайлика

==<{P<=<{P{=<{PP

=<{=<{P

=<{

==<{P

<=<{P

{

PPP

должно быть 6 смайликов

Исходный Код

```
import re

filename: str = input("Enter the name of the file to read: ")
file = open(filename, "r")
text = file.read()
print("\nText from file \' + filename + "\':\n" + text)

match = r"=<{P"
matches = re.findall(match, text)
count = len(matches)
print("The amount of times =<{P appeared: ", count)
```

Результат

Первый тест:

Enter the name of the file to read: sm1

Text from file 'sm1':

=</**/{P" aaaaaaaaaaaaaaaaaa

aa

=<{P

=<{P" cvx ={"

={P"aaaaaaa<aaa =<{P""=<{P"=<{P

должно быть 5 смайликов

The amount of times =<{P appeared: 5

Второй тест:

Enter the name of the file to read: sm2

Text from file 'sm2':

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Sed felis eget velit aliquet. In ornare quam viverra orci sagittis eu volutpat odio facilisis. Lacus viverra vitae congue eu. Risus at ultrices mi tempus imperdiet nulla malesuada. Condimentum id venenatis a condimentum. Tellus in metus vulputate eu scelerisque felis imperdiet. Massa vitae tortor condimentum lacinia quis vel eros donec. Aliquam eleifend mi in nulla posuere sollicitudin. Faucibus interdum posuere lorem ipsum dolor. cursus eget nunc scelerisque viverra mauris in aliquam.

должно быть 0 смайлика

The amount of times =<{P appeared: 0

Третий тест:

Enter the name of the file to read: sm3

Text from file 'sm3':

'=<{P'""=<{P" =<{P

=

<

{

P

=

<{

P

= <{ P

должно быть 3 смайлика

The amount of times =<{P appeared: 3

Четвёртый тест:

Enter the name of the file to read: sm4

Text from file 'sm4':

должно быть 0 смайлика

The amount of times =<{P appeared: 0

Пятый тест:

Enter the name of the file to read: sm5

Text from file 'sm5':

==<{P<=<{P{=<{PP

=<{=<{P

=<{

==<{P

<=<{P

{

PPP

должно быть 6 смайликов

The amount of times =<{P appeared: 6

Доп. задание №1

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.

Тесты

Файл 'sn1':

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

```
imcx xnm
```

Файл 'sn2':

Lorem dolor sed viverra ipsum nunc. Diam vel quam elementum pulvinar etiam non. Integer vitae justo eget magna. Vitae suscipit tellus mauris a. Id velit ut tortor pretium viverra suspendisse potenti. Pellentesque habitant morbi tristique senectus et netus et malesuada fames. Sit amet risus nullam eget felis eget nunc. Volutpat blandit aliquam etiam erat velit scelerisque. Semper viverra nam libero justo laoreet. Hjkpd P.Q. Non enim praesent elementum facilisis leo vel fringilla est. Viverra tellus in hac habitasse platea dictumst. Commodu ullamcorper a lacus vestibulum sed. Rutrum quisque non tellus orci ac auctor augue mauris augue. Accumsan tortor posuere ac ut consequat semper viverra nam libero. Vel facilisis volutpat est velit egestas. Quisque sagittis purus sit amet volutpat consequat. Erat imperdiet sed euismod nisi porta lorem. At volutpat diam ut venenatis tellus in metus.

Файл 'sn3':

Anitschenko A.A.Анищенко А.А.

Anitschenko A.A.Анищенко А.А.Anitschenko A.A.Анищенко А.А.

Файл 'sn4':

Машина

Е.А.Машина Е.А.Машина Е.А.Машина

Е.А.Машина Е.А.

Файл 'sn5':

При адресовании должностному или физическому лицу инициалы указывают перед фамилией: И. И. Иванову, А. А. Петрову. При составлении библиографии или алфавитных списков лиц инициалы пишутся после фамилии, поскольку поиск рубрики указателя ведется по буквам фамилии: Иванов И. И.

Исходный Код

```
import re

filename: str = input("Enter the name of the file to read: ")
file = open(filename, "r")
text = file.read()
print("\nText from file \'' + filename + '\':\n" + text)
```

```
match = r" ([А-ЯА-З][\w-]*)\s*[А-ЯА-З][\.\s]*[А-ЯА-З]\."
surname_list = re.findall(match, text)
surname_list.sort()
print("List of surnames: \n")
for x in range(len(surname_list)):
    print(surname_list[x])
```

Результат

Первый тест:

Enter the name of the file to read: sn1

Text from file 'sn1':

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

imcx xnm

List of surnames:

Анищенко

Балакшин

Машина

Второй тест:

Enter the name of the file to read: sn2

Text from file 'sn2':

Lorem dolor sed viverra ipsum nunc. Diam vel quam elementum pulvinar etiam non. Integer vitae justo eget magna. Vitae suscipit tellus mauris a. Id velit ut tortor pretium viverra suspendisse potenti. Pellentesque habitant morbi tristique senectus et netus et malesuada fames. Sit amet risus nullam eget felis eget nunc. Volutpat blandit aliquam etiam erat velit scelerisque. Semper viverra nam libero justo laoreet. Hjkpd P.Q. Non enim praesent elementum facilisis leo vel fringilla est. Viverra tellus in hac habitasse platea dictumst. Commodo ullamcorper a lacus vestibulum sed. Rutrum quisque non tellus orci ac auctor augue mauris augue. Accumsan tortor posuere ac ut consequat semper viverra nam libero. Vel facilisis volutpat est velit egestas. Quisque sagittis purus sit amet volutpat consequat. Erat imperdiet sed euismod nisi porta lorem. At volutpat diam ut venenatis tellus in metus.

List of surnames:

Hjkpd

Третий тест:

Enter the name of the file to read: sn3

Text from file 'sn3':

Anitschenko A.A.Анищенко А.А.

Anitschenko A.A.Анищенко А.А.Anitschenko A.A.Анищенко А.А.

List of surnames:

Anitschenko

Anitschenko

Anitschenko

Анищенко

Анищенко

Анищенко

Четвёртый тест:

Enter the name of the file to read: sn4

Text from file 'sn4':

Машина

Е.А.Машина Е.А.Машина Е.А.Машина

Е.А.Машина Е.А.

List of surnames:

Машина

Машина

Машина

Машина

Машина

Пятый тест:

Enter the name of the file to read: sn5

Text from file 'sn5':

При составлении библиографии или алфавитных списков лиц инициалы пишутся после фамилии, поскольку поиск рубрики указателя ведется по буквам фамилии:
Иванов И. И.

List of surnames:

ИВАНОВ

Доп. задание №2

Вывесили списки стипендиатов текущего семестра, которые представляют из себя список людей ФИО и номер группы этого человека. Вы решили подшутить над некоторыми из своих одногруппников и удалить их из списка.

С помощью регулярного выражения найдите всех студентов своей группы, у которых инициалы начинаются на одну и ту же букву и исключите их из списка.

Тесты

Файл 'pr1':

Петров П.П. P000

Анищенко А.А. P33113

Примеров Е.В. P000

Иванов И.И. P000

Файл 'pr2':

Petrov P.P. P000

Анищенко А.А. P33113

Primerov E.V. P000

Ivanov I.I. P000

Файл 'pr3':

Петров П.П. P000

Иванов И.И. P000

Анищенко А.А. P33113

Примеров Е.В. P000

Аванов А.А. P000

Иванов И.И. P1000

Иванов И.И. P000

Петров-Иванов П.П. P000

Файл 'pr4':

Вайкус Я.М. P3124

Владимиров В.А. P3124

Гафурова Ф.Ф. P3124

Грибич Е.Д. P3124

Клименко В.И. P3124

Кобик Н.А. P3124

Колесников А.И. P3124

Кононыхин К.Д. P3124

Крохин Р.О. P3124

Круглов Е.И. P3124

Леонов Р.В. P3124

Маликов Г.И. P3124

Хвостова И.Л. P3124

Черневская К.А. P3124

Файл 'pr5':

Я Я.Я. P000

В В.В. P3124

Ф Ф.Ф. P000

Исходный Код

```
import re

filename: str = input("Enter the name of the file to read: ")
file = open(filename, "r")
text = file.read()
group: str = input("Enter your group: ")
print("\nText from file \' + filename + "\':\n" + text)

line = text.split("\n")
line = list(filter(None, line)) #removes empty strings from list
print(line)
S_re = r"^\s*([A-ЯА-З])"
I_re = r"\s([A-ЯА-З])\."
O_re = r"\.([A-ЯА-З])"
G_re = r"P\d+"
S_lst = re.findall(S_re, text)
print(S_lst)
I_lst = re.findall(I_re, text)
print(I_lst)
O_lst = re.findall(O_re, text)
print(O_lst)
G_lst = re.findall(G_re, text)
print(G_lst)

print("Updated list: \n")
for i in range(len(S_lst)):
    if not (S_lst[i] == I_lst[i] == O_lst[i] and G_lst[i] == group):
        print(line[i])
```

Результат

Первый тест:

Enter the name of the file to read: pr1

Enter your group: P000

Text from file 'pr1':

Петров П.П. P000

Анищенко А.А. P33113

Примеров Е.В. P000

Иванов И.И. P000

```
['Петров П.П. P000', 'Анищенко А.А. P33113', 'Примеров Е.В. P000', 'Иванов И.И. P000']
```

```
['П', 'А', 'П', 'И']
```

```
['П', 'А', 'Е', 'И']
```

```
['П', 'А', 'В', 'И']
```

```
['P000', 'P33113', 'P000', 'P000']
```

Updated list:

Анищенко А.А. P33113

Примеров Е.В. P000

Второй тест:

Enter the name of the file to read: pr2

Enter your group: P000

Text from file 'pr2':

Petrov P.P. P000

Анищенко А.А. P33113

Primerov E.V. P000

Ivanov I.I. P000

```
['Petrov P.P. P000', 'Анищенко А.А. P33113', 'Primerov E.V. P000', 'Ivanov I.I. P000']
```

```
['P', 'А', 'P', 'I']
```

```
['P', 'А', 'Е', 'I']
```

```
['P', 'А', 'V', 'I']
```

```
['P000', 'P33113', 'P000', 'P000']
```

Updated list:

Анищенко А.А. P33113

Primerov E.V. P000

Третий тест:

Enter the name of the file to read: pr3

Enter your group: P000

Text from file 'pr3':

Петров П.П. P000

Иванов И.И. P000

Анищенко А.А. P33113

Примеров Е.В. P000

Аванов А.А. P000

Иванов И.И. P1000

Иванов И.И. P000

Петров-Иванов П.П. P000

```
['Петров П.П. P000', 'Иванов И.И. P000', 'Анищенко А.А. P33113',  
'Примеров Е.В. P000', 'Аванов А.А. P000', 'Иванов И.И. P1000', 'Иванов И.И.  
P000', 'Петров-Иванов П.П. P000']
```

```
['П', 'И', 'А', 'П', 'А', 'И', 'И', 'П']
```

```
['П', 'И', 'А', 'Е', 'А', 'И', 'И', 'П']
```

```
['П', 'И', 'А', 'В', 'А', 'И', 'И', 'П']
```

```
['P000', 'P000', 'P33113', 'P000', 'P000', 'P1000', 'P000', 'P000']
```

Updated list:

Анищенко А.А. P33113

Примеров Е.В. P000

Иванов И.И. P1000

Четвёртый тест:

Enter the name of the file to read: pr4

Enter your group: P3124

Text from file 'pr4':

Вайкус Я.М. P3124

Владимиров В.А. P3124

Гафурова Ф.Ф. P3124

Грибич Е.Д. P3124

Клименко В.И. P3124

Кобик Н.А. P3124

Колесников А.И. P3124

Кононыхин К.Д. P3124

Крохин Р.О. P3124

Круглов Е.И. P3124

Леонов Р.В. P3124

Маликов Г.И. P3124

Хвостова И.Л. P3124

Черневская К.А. P3124

['Вайкус Я.М. P3124', 'Владимиров В.А. P3124', 'Гафурова Ф.Ф. P3124', 'Грибич Е.Д. P3124', 'Клименко В.И. P3124', 'Кобик Н.А. P3124', 'Колесников А.И. P3124', 'Кононыхин К.Д. P3124', 'Крохин Р.О. P3124', 'Круглов Е.И. P3124', 'Леонов Р.В. P3124', 'Маликов Г.И. P3124', 'Хвостова И.Л. P3124', 'Черневская К.А. P3124']

['В', 'В', 'Г', 'Г', 'К', 'К', 'К', 'К', 'К', 'К', 'Л', 'М', 'Х', 'Ч']

['Я', 'В', 'Ф', 'Е', 'В', 'Н', 'А', 'К', 'Р', 'Е', 'Р', 'Г', 'И', 'К']

['М', 'А', 'Ф', 'Д', 'И', 'А', 'И', 'Д', 'О', 'И', 'В', 'И', 'Л', 'А']

['P3124', 'P3124', 'P3124', 'P3124', 'P3124', 'P3124', 'P3124', 'P3124', 'P3124', 'P3124', 'P3124', 'P3124', 'P3124']

Updated list:

Вайкус Я.М. P3124

Владимиров В.А. P3124

Гафурова Ф.Ф. P3124

Грибич Е.Д. P3124

Клименко В.И. P3124

Кобик Н.А. P3124

Колесников А.И. P3124

Кононыхин К.Д. P3124

Крохин Р.О. P3124

Круглов Е.И. P3124

Леонов Р.В. P3124

Маликов Г.И. P3124

Хвостова И.Л. P3124

Черневская К.А. P3124

Пятый тест:

Enter the name of the file to read: pr5

Enter your group: P3124

Text from file 'pr5':

Я Я.Я. P000

В В.В. P3124

Ф Ф.Ф. P000

['Я Я.Я. P000', 'В В.В. P3124', 'Ф Ф.Ф. P000']

['Я', 'В', 'Ф']

['Я', 'В', 'Ф']

['Я', 'В', 'Ф']

['P000', 'P3124', 'P000']

Updated list:

Я Я.Я. P000

Ф Ф.Ф. P000

Process finished with exit code 0

Вывод

В данной работе были использованы регулярные выражения на языке Python с помощью модуля `re`. Регулярные выражения позволяют быстро произвести такие операции над текстами как поиск и замена. Преимущество регулярных выражений в том, что они могут быть использованы также в других языках программирования.

Список Литературы

Python Software Foundation (2022) re — Regular expression operations:
<https://docs.python.org/3/library/re.html>

Goyvaerts J. (2021) Regular Expressions: <http://www.regular-expressions.info/tutorial.html>