

Национальный исследовательский университет информационных технологий,
механики и оптики

Факультет программной инженерии и компьютерной техники

Лабораторная работа №4 «Регулярные выражения»

Работу выполнил:

Бавыкин Роман

Преподаватель:

Балакшин Павел Валерьевич

Группа: Р3110

Номер в группе: 4

Санкт-Петербург

2020 г.

Задание на 70 баллов

- 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.

Вариант = $4 \% 5 = 4$

Анатолий выложил пост с расписанием доп. занятий по информатике, но везде перепутал время. Поэтому нужно заменить все вхождения времени на строку (TBD).

Время – это строка вида HH:MM:SS или HH:MM, в которой HH – число от 00 до 23, а MM и SS – число от 00 до 59.

Исходный код программы:

```
1 import re
2 for i in range(5):
3     fileName = 'test' + str(i+1) + '.txt'
4     file = open(fileName, encoding="utf-8")
5     text = file.read()
6     print(text + '\n')
7     pattern = re.compile(r"\b(?:[0-1]\d|2[0-3])(?:[0-5]\d){1,2}(?=\W)")
8     matches = pattern.findall(text)
9     counter = 0
10    for time in matches:
11        print('Change ' + time + ' to ', end='')
12        matches[counter] = input()
13        counter += 1
14    for time in reversed(matches):
15        text = pattern.sub(matches[counter - 1], text, counter)
16        counter -= 1
17    print('\n' + text + '\n')
18    file.close()
```

Тесты:

1. Уважаемые студенты! В эту субботу в, 15:00 планируется доп. занятие на 2 часа. То есть в 17:00:01 оно уже точно кончится.
2. 21:59:56, 21:59:57, 21:59:58. Время шло очень медленно, а я не мог дождаться, когда уже наступит 22:00.
3. 02.09.2020 у группы Р3110 будет два, лабораторных занятия по информатике:
1 - с 11:40 до 13:10;

2 - с 13:30 до 15:00.

4. В армии подъем состава в 07:00-07:10, а отбой в 23:00.
5. Если студент отчислится, то, у него не будет пар в 08:00, и он сможет вставать не в 07:00, а в 11:00 или вообще в 15:00

Результат работы программы:

Уважаемые студенты! В эту субботу в, 15:00 планируется доп. занятие на 2 часа. То есть в 17:00:01 оно уже точно кончится.

Change 15:00 to 16:00
Change 17:00:01 to 18:00:01

Уважаемые студенты! В эту субботу в, 16:00 планируется доп. занятие на 2 часа. То есть в 18:00:01 оно уже точно кончится.

21:59:56, 21:59:57, 21:59:58. Время шло очень медленно, а я не мог дожидаться, когда уже наступит 22:00.

Change 21:59:56 to 21:57
Change 21:59:57 to 21:58
Change 21:59:58 to 21:59
Change 22:00 to 22:00

21:57, 21:58, 21:59. Время шло очень медленно, а я не мог дожидаться, когда уже наступит 22:00.

02.09.2020 у группы Р3110 будет два, лабораторных занятия по информатике:
1 - с 11:40 до 13:10;
2 - с 13:30 до 15:00.

Change 11:40 to 08:20
Change 13:10 to 09:50
Change 13:30 to 10:00
Change 15:00 to 11:30

02.09.2020 у группы Р3110 будет два, лабораторных занятия по информатике:
1 - с 08:20 до 09:50;
2 - с 10:00 до 11:30.

В армии подъем состава в 07:00-07:10,, а отбой в 23:00.

Change 07:00 to 07:30
Change 07:10 to 07:40
Change 23:00 to 23:30

В армии подъем состава в 07:30-07:40,,
а отбой в 23:30.

Если студент отчислится, то,
у него не будет пар в 08:00,
и он сможет вставать не в 07:00,
а в 11:00 или вообще в 15:00.

Change 08:00 to 08:20
Change 07:00 to 06:00
Change 11:00 to 12:00
Change 15:00 to 14:00

Если студент отчислится, то,
у него не будет пар в 08:20,
и он сможет вставать не в 06:00,
а в 12:00 или вообще в 14:00.

Результаты работы программы корректны и совпадают с ответами, найденными вручную по данным тестам.

Доп. задание No1 (+12 баллов)

1) Для выполнения задания скачайте текстовый документ по ссылке, соответствующей вашему варианту .

Вариант = $4 \% 30 = 4$

Гамлет: <https://drive.google.com/file/d/1j2p-EAwainLQ7dbOf2CFAXDTjQ9O7tNI/view>

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

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

Вариант = $4 \% 4 = 0$

Все предложения, в которых ровно 6 слов, одно из которых двусложное.

Исходный код программы:

```
1 import re
2 file = open('Hamlet.txt', encoding="utf-8")
3 text = file.read()
4 pattern = re.compile(r"(?<=[.!?])\s[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[eEyYuUiIoOaA]
[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[eEyYuUiIoOaA]
[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*(?:[.,:;'\"]\s|)]+(?:
[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[eEyYuUiIoOaA]
[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*)"
```

[illegible]

	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[eEyYuUiloOaA]
	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*
	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[eEyYuUiloOaA]
	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[eEyYuUiloOaA]
	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[eEyYuUiloOaA]
	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*(?:
	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[eEyYuUiloOaA]*)[\\\"\\'])*[.?!](?<=[.!?]\\s)
	(?:(?:[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[eEyYuUiloOaA]
	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*)
	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[eEyYuUiloOaA]
	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[eEyYuUiloOaA]
	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[eEyYuUiloOaA]
	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*(?:
	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[eEyYuUiloOaA]*)?(?:[\\.,:;\\\"\\'\\s\\])+){5}
	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[eEyYuUiloOaA]
	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[eEyYuUiloOaA]
	[qQwWrRtTpPsSdDfFgGhHjJkKlLzZxXcCvVbBnNmM]*[\\\"\\'])*[.?!]"
5	for sentence in pattern.findall(text):
6	print (sentence + '\\n')
7	file.close()

Результат работы программы:

Is it not like the King?
Nay, not so much, not two.
My lord, the King your father.
Did you not speak to it?
Then saw you not his face?
I think it lacks of twelve.
In faith,
My lord, not I.
We have sworn, my lord, already.
My lord, I did intend it.
My lord, that would dishonour him.
And let him ply his music.
I will go seek the King.
Come, go we to the King.
That do I long to hear.
He said I was a fishmonger.
What is the matter, my lord?
In the secret parts of Fortune?

But your news is not true.

Let me
question more in particular.

To me it is a prison.

My lord, we were sent for.

in apprehension how like a god!

Gentlemen, you are welcome to Elsinore.

Go thy ways to a nunnery.

O, help him, you sweet heavens!

We shall know by this fellow.

What do you call the play?

Upon the talk of the poisoning?

It is as easy as lying.

No, by the rood, not so!

A king of shreds and patches!

How is it with you, lady?

To whom do you speak this?

The nephew to old Norway, Fortinbras.

I will not speak with her.

How long hath she been thus?

I hope all will be well.

Do you see this, O God?

What out of this, my lord?

How dost thou understand the Scripture?

Ay, tell me that, and unyoke.

Not a jot more, my lord.

Whose do you think it was?

Prithee, Horatio,
tell me one thing.

What wilt thou do for her?

For love of God, forbear him!

A hit, a very palpable hit.

What is it you will see?

All this can I
Truly deliver.

Для проверки корректности работы программы искал нужные предложения в небольшой части текста и сверял с тем, что выводит программа. Результат работы корректный.

Доп. задание No2 (+18 баллов)

- 1) Реализуйте программный продукт на языке Python, используя регулярные выражения по варианту, представленному в таблице.
- 2) Для своей программы придумайте минимум 5 тестов.
- 3) Протестируйте свою программу на этих тестах.

Вариант = $4 \% 6 = 4$

С помощью регулярного выражения найти в тексте все слова, в которых две гласные стоят подряд, а после этого слова идёт слово, в котором не больше 3 согласных.

Исходный код программы:

```
1 import re
2 for i in range(5):
3     fileName = 'addtest' + str(i+1) + '.txt'
4     file = open(fileName, encoding="utf-8")
5     text = file.read()
6     print(text + '\n')
7     pattern = re.compile(r"\b\w*[aAoOyУэЭыЫюЮяЯеЕиИ]{2}\w*\b(?:=[.,\[\]\'\";:\  
s]*[aAoOyУэЭыЫюЮяЯеЕиИть]*[БбВвГгДдЖжЗзКкЛлМмНнПпРрСсТтФфХхЦцЧчШшЩщ]  
[aAoOyУэЭыЫюЮяЯеЕиИть]*\b|[\.\[\]\'\";:\  
s]*[aAoOyУэЭыЫюЮяЯеЕиИть]*[БбВвГгДдЖжЗзКкЛлМмНнПпРрСсТтФфХхЦцЧчШшЩщ]  
[aAoOyУэЭыЫюЮяЯеЕиИть]*[БбВвГгДдЖжЗзКкЛлМмНнПпРрСсТтФфХхЦцЧчШшЩщ]  
[aAoOyУэЭыЫюЮяЯеЕь]*\b|[\.\[\]\'\";:\  
s]*[aAoOyУэЭыЫюЮяЯеЕь]*[БбВвГгДдЖжЗзКкЛлМмНнПпРрСсТтФфХхЦцЧчШшЩщ]  
[aAoOyУэЭыЫюЮяЯеЕиИть]*[БбВвГгДдЖжЗзКкЛлМмНнПпРрСсТтФфХхЦцЧчШшЩщ]  
[aAoOyУэЭыЫюЮяЯеЕиИть]*[БбВвГгДдЖжЗзКкЛлМмНнПпРрСсТтФфХхЦцЧчШшЩщ]  
[aAoOyУэЭыЫюЮяЯеЕиИть]*\b)")
8     matches = pattern.findall(text)
9     for word in matches:
10         print(word)
11     print()
12     file.close()
```

Тесты:

1. Кривошеее существо гуляет по парку.
2. Сегодня на лекции Павел Валерьевич учил нас тому, как быстро написать много ПСЖ.
3. Удвоенные гласные, но только удвоенные перед словами с не более чем тремя согласными.
4. Мы с друзьями хотим написать ПСЖ и пойти в армию, потому что в армии весело.
5. Нужно хорошо подготовиться к защите лабораторной работе по информатике, чтобы сдать ее на 100 баллов.

Результат работы программы:

Кривошеее существо гуляет по парку.

гуляет

Сегодня на лекции Павел Валерьевич учил нас
тому, как быстро написать много ПСЖ.

лекции

Удвоенные гласные, но только удвоенные перед словами
с не более чем тремя согласными.

гласные
удвоенные
более

Мы с друзьями хотим написать ПСЖ
и пойти в армию, потому что в армии весело.

армию
армии

Нужно хорошо подготовиться к выполнению лабораторной
работе по информатике, чтобы сдать ее на 100 баллов.

ее

Вывод: во время выполнения лабораторной работы научился составлять регулярные выражение и использовать их в языке Python 3.X.