

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО

Факультет Программной инженерии и компьютерной техники

Информатика

Лабораторная работа № 3 "Регулярные выражения и языки разметки
документов. Python"

Выполнил студент

Егорова Варвара Александровна

Группа № Р3123

Преподаватель: Болдырева Елена Александровна

г. Санкт-Петербург

2023

Оглавление

Задание №1:.....	3
(Таблица 1)	3
Задание №2:.....	5
Задание №3:.....	6
Дополнительное задание №2:	9
Вывод.....	11
Список литературы	12

Задание №1:

Необходимо реализовать программный продукт на Python, используя регулярные выражения по варианту, представленному в таблице 1. Для своей программы нужно придумать 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. Программа должна считать количество смайликов определённого вида в предложенном тексте. Все смайлики имеют структуру [глаза][нос][рот].

Номер в ИСУ %5	Глаза	Номер в ИСУ %4	Нос	Номер в ИСУ %7	Рот
0	:	0	-	0	(
1	;	1	<	1)
2	X	2	-{	2	O
3	8	3	<{	3	
4	=			4	\
				5	/
				6	P

(Таблица 1)

Исходный код:

```

# Вариант 036
import re
import sys

sys.stdin = open('test_1.txt')

def smile(isu):
    eyes = [':', ';', 'X', ';', '8', '=']
    nose = ['-', '<', '-{', '<{']
    mouth = ['(', ')', 'O', '|', '\\', '/', 'P']
    sml = eyes[isu % 5] + nose[isu % 4] + mouth[isu % 7]
    return (sml)

def ammount(string, patt):
    n = len(re.findall(patt, string))
    return (n)

patt = smile(int(input()))
print('Ваш смайл: ', patt)
for _ in range(5):
    string = input()
    print('Исходная строка: ', string)
    print('Количество смайлов: ', end='')
    print(ammount(string, patt))
    print('')

```

Входной файл с пятью тестами:

```

408575
:<0|\=8=-{):<{P()|\ /X-):<{P:<{P:<{P
8X:<{P{P-<0:<|8-):<{:<{PP:-{\:<{P
X<{/P=-//:<{P:<{P:<{P0\|:-0:<{P
::=--{(=-|:<{PX:</=<{PX<):<{P8-{)
=<0:-{X<{|:-PX<08-{\\:-/;-P;-{-|8-/

```

При подсчете смайликов вручную:

В первом тесте: 4

Во втором: 3

В третьем: 4

В четвертом: 2

В пятом: 0

Результат работы программы:

```

Ваш смайл: :<{P
Исходная строка: :<0|\=8=-{):<{P()|\/X-):<{P:<{P:<{P
Количество смайлов: 4

Исходная строка: 8X:<{P{P-<0:<|8-):<{:<{PP:-{\:<{P
Количество смайлов: 3

Исходная строка: X<{/P=-//:<{P:<{P:<{P0\|:-0:<{P
Количество смайлов: 4

Исходная строка: ::=---{(-=|:<{PX:</=<{PX<):<{P8-{)
Количество смайлов: 2

Исходная строка: =<0:-{X<{|:-PX<08-{\: -/;-P:;-{|8-/
Количество смайлов: 0

```

Задание №2:

Необходимо реализовать программный продукт на языке Python, используя регулярные выражения (по варианту), придумать минимум 5 тестов и протестировать свою программу на этих тестах.

Вариант 5:

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 — числа от 00 до 59.

Исходный код:

```

import re
import sys

sys.stdin = open('test_2.txt')

def time(text):
    patt = "[0-2][0-23]:[0-5][0-9](:[0-5][0-9])?|[0-1][0-9]:[0-5][0-9](:[0-5][0-9])?"
    text = re.sub(patt, '(TBD)', text)
    return text

for _ in range(5):
    text = input()
    print('Исходный текст: ', text)
    print('Полученный текст: ', end = '')
    print(time(text))
    print('')

```

Тесты:

1. Dear students! Additional service is planned this Saturday at 15:00. lesson for 2 hours. That is, at 17:00:01 it will definitely end.
2. Dear students! Additional service is planned this Saturday at 15:60. lesson for 2 hours. That is, at 24:17:01 it will definitely end.
3. Dear! If you do not return the suitcase by 09:00, then at 09:00:01 I will not be responsible for myself. PS. With a ratio of 25:50 everything is fine! 23:55:32:33
4. Dear students and staff! We inform you that on 26.09 from 12:07:01 to 23:59 the canteen will be closed!
5. For earthlings, time looks like: 22:01 or 12:01:04 (24 hours in a day), and on Mars we have 32 hours in a day, so time looks like: 32:54:11 or 27:32.

Результат работы программы:

```
Исходный текст: Dear students! Additional service is planned this Saturday at 15:00. lesson for 2 hours. That is, at 17:00:01 it will definitely end.
Полученный текст: Dear students! Additional service is planned this Saturday at (TDB). lesson for 2 hours. That is, at (TDB) it will definitely end.

Исходный текст: Dear students! Additional service is planned this Saturday at 15:60. lesson for 2 hours. That is, at 24:17:01 it will definitely end.
Полученный текст: Dear students! Additional service is planned this Saturday at 15:60. lesson for 2 hours. That is, at 24:(TDB) it will definitely end.

Исходный текст: Dear! If you do not return the suitcase by 09:00, then at 09:00:01 I will not be responsible for myself. PS. With a ratio of 25:50 everything is fine! 23:55:32
Полученный текст: Dear! If you do not return the suitcase by (TDB), then at (TDB) I will not be responsible for myself. PS. With a ratio of 25:50 everything is fine! (TDB):33

Исходный текст: Dear students and staff! We inform you that on 26.09 from 12:07:01 to 23:59 the canteen will be closed!
Полученный текст: Dear students and staff! We inform you that on 26.09 from (TDB) to (TDB) the canteen will be closed!

Исходный текст: For earthlings, time looks like: 22:01 or 12:01:04 (24 hours in a day), and on Mars we have 32 hours in a day, so time looks like: 32:54:11 or 27:32.
Полученный текст: For earthlings, time looks like: (TDB) or (TDB) (24 hours in a day), and on Mars we have 32 hours in a day, so time looks like: 32:54:11 or 27:32.
```

Задание №3:

Необходимо определить номер варианта как остаток деления номера в ИСУ на 36. Исходя из структуры расписания, сформировать файл с расписанием в формате, указанном в задании в качестве исходного. Написать программу на языке Python 3.x, которая бы осуществляла парсинг и конвертацию исходного файла в новый. Нельзя использовать готовые библиотеки, в том числе регулярные выражения в Python и библиотеки для загрузки XML-файлов.

Вариант 11:

Исходный формат — XML, Результирующий формат — YAML.

Исходное расписание в формате XML:

```
<schedule>
  <Monday>
    <ComputerScience>
      <time>8:20-9:50</time>
      <teacher>Boldyreva Elena Aleksandrovna</teacher>
      <place>Online</place>
    </ComputerScience>
    <ComputerScience>
      <time>10:00-11:30</time>
      <teacher>Boldyreva Elena Aleksandrovna</teacher>
      <place>Online</place>
    </ComputerScience>
    <English>
      <time>11:40-13:10</time>
      <teacher>Abdullina Nailya Zufarovna</teacher>
      <place>Lomonosova 9</place>
      <classroom>3225</classroom>
    </English>
    <English>
      <time>13:30-15:00</time>
      <teacher>Abdullina Nailya Zufarovna</teacher>
      <place>Lomonosova 9</place>
      <classroom>3225</classroom>
    </English>
  </Monday>
  <Tuesday>
  <Wednesday>
  <Thursday>
  <Friday>
  <Saturday>
</schedule>
```

Исходный код программы:

```
xml_file = open('C://Users//Egorova Varvara//Desktop//schedule.xml', 'r')
yaml_file = open('out.txt', 'w')

flag = 0
counter = 0
for line in xml_file:
    xml_string = line
    yaml_string = ''
    for i in range(len(xml_string)):
        symbol = xml_string[i]
        if symbol == '\\t':
            yaml_string += '\\t'
        elif symbol == ' ':
            yaml_string += ' '
        elif symbol == '<':
            if xml_string[i + 1] != '/':
                if flag != 0:
                    yaml_string += '-'
                else:
                    flag = 1
                    continue
            else:
                break
        elif symbol.isalpha() or symbol.isdigit() or symbol in ':-':
            yaml_string += symbol
        elif symbol == '>':
            yaml_string += ': '
            if i + 1 == len(xml_string) - 1:
                counter += 1
    num = yaml_string.count('\\t')
    if num * '\\t' != yaml_string:
        yaml_file.write(yaml_string + '\\n')
```


Результат работы программы:

```
schedule:
  -Monday:
    -ComputerScience:
      -time: 8:20-9:50
      -teacher: Boldyreva Elena Aleksandrovna
      -place: Online
    -ComputerScience:
      -time: 10:00-11:30
      -teacher: Boldyreva Elena Aleksandrovna
      -place: Online
    -English:
      -time: 11:40-13:10
      -teacher: Abdullina Nailya Zufarovna
      -place: Lomonosova 9
      -classroom: 3225
    -English:
      -time: 13:30-15:00
      -teacher: Abdullina Nailya Zufarovna
      -place: Lomonosova 9
      -classroom: 3225
  -Tuesday:
    -FundamentalsOfDiscreteMath:
      -time: 10:00-11:30
      -teacher: Lisicyna Lybov Sergeevna
      -place: Kronverksky avenue 49
      -classroom: 24357
    -MathAnalys:
      -time: 13:30-15:00
      -teacher: Pravdin Konstantyn Vladymirovich
      -place: Online
    -MathAnalys:
      -time: 15:20-16:50
      -teacher: Isaeva Tatyana Tinofeevna
      -place: Online
  -Wednesday:
  -Thursday:
  -Friday:
  -Saturday:
```

Дополнительное задание №2:

Переписать исходный код, добавив в него использование регулярных выражений. Сравнить полученные результаты и объяснить их сходство/различие.

Исходный код программы:

```
import re
xml_file = open('C://Users//Egorova Varvara//Desktop//schedule.xml', 'r')
yaml_file = open('out.txt', 'w')

flag = 0
counter = 0
for line in xml_file:
    xml_string = line
    yaml_string = ''
    patt = re.findall(r'\s{0,}<\w{1,}>', xml_string)
    if patt:
        name = patt[0]
        name = name.replace('<', '')
        name = name.replace('>', '')
        if name[0] == '\t':
            num = name.count('\t')
            if num != 0:
                name = num * '\t' + '-' + name[num:]
        yaml_string += name + ':'
    patt = re.findall(r'>(.*?)<', xml_string)
    yaml_string += (patt[0] if patt else '')
    yaml_file.write(yaml_string + '\n' if yaml_string else '')
```

Результат работы программы:

```
schedule:
  -Monday:
    -ComputerScience:
      -time: 8:20-9:50
      -teacher: Boldyreva Elena Aleksandrovna
      -place: Online
    -ComputerScience:
      -time: 10:00-11:30
      -teacher: Boldyreva Elena Aleksandrovna
      -place: Online
    -English:
      -time: 11:40-13:10
      -teacher: Abdullina Nailya Zufarovna
      -place: Lomonosova 9
      -classroom: 3225
    -English:
      -time: 13:30-15:00
      -teacher: Abdullina Nailya Zufarovna
      -place: Lomonosova 9
      -classroom: 3225
  -Tuesday:
    -FundamentalsOfDiscreteMath:
      -time: 10:00-11:30
      -teacher: Lisicyna Lybov Sergeevna
      -place: Kronverksky avenue 49
      -classroom: 24357
    -MathAnalys:
      -time: 13:30-15:00
      -teacher: Pravdin Konstantyn Vladymirovich
      -place: Online
    -MathAnalys:
      -time: 15:20-16:50
      -teacher: Isaeva Tatyana Tinofeevna
      -place: Online
  -Wednesday:
  -Thursday:
  -Friday:
  -Saturday:
```

Различий в результате работы обоих кодов нет, однако первый код выполняется дольше из-за наличия вложенных циклов.

Вывод

В ходе выполнения данной лабораторной работы я изучила регулярные выражения, научилась работать с ними в Python, изучила основы XML и YAML и научилась делать парсинг из одного формата в другой.

Список литературы

1. Регулярные выражения в Python от простого к сложному. Подробности, примеры, картинки, упражнения // habr.com: блог об IT. 2007 URL: <https://habr.com/ru/post/349860/>
2. XML and JSON. // habr.com: блог об IT. 2007 URL: https://habr.com/ru/company/rambler_and_co/blog/525498/