

Практика 3. Регулярные выражения

Работа выполняется при помощи Python.

Проверить регулярное выражение можно по ссылке <https://regex101.com/>

У каждой строки есть начало и конец. Начинается строка с символа `^`, заканчивается `$`.

Диапазон обозначается в квадратных скобках `[]`.

Диапазон цифр от 0 до 9 и строчных букв от а до z будет выглядеть как `[0-9a-z]`.

Диапазон зависит от регистра, если нужно добавить так же заглавные буквы, получится диапазон вида `[0-9-a-zA-Z]`.

Для поиска при помощи регулярных выражений в Python используется модуль `re`.

`re.findall` находит все совпадения в соответствии с выражением.

```
import re

data = 'какой-то текст 123' # Текст, по которому идёт поиск
reg = r'[0-9]' # Регулярное выражение для поиска

re.findall(reg, data)
>>> ['1', '2', '3']
```

Регулярное выражение искало отдельный символ и нашло 3 соответствия.

Чаще нам необходимо искать набор символов.

Например мы можем указать кол-во символов для поиска:

```
re.findall(r'[0-9]{2}', 'какой-то текст 123')
>>> ['12']
```

Можно указать диапазон. Можно указать только начало или конец диапазона.

```
data = ['a12', 'a12345', 'a12345678']
result_1 = []
result_2 = []
result_3 = []

for i in data:
    result_1.append(re.findall(r'[0-9]{3,6}', i))
    result_2.append(re.findall(r'[0-9]{3}', i))
    result_3.append(re.findall(r'[0-9]{3,}', i))

result_1
>>> [[], ['12345'], ['123456']]
```

```
result_2
>>> [[], ['123'], ['123', '456']]

result_3
>>> [[], ['12345'], ['12345678']]
```

У диапазонов есть синонимы (кванторы):

```
? - {0,1}
* = {0,}
+ = {1,}
```

В регулярное выражение можно записать любой символ. Если он является служебным, то его можно экранировать. `\(\\)`

Так же есть уже готовые наборы символов:

```
. - любой символ кроме \n
\d - цифра
\D - всё, кроме цифры
\s - пробел, табуляция, \n
\S - всё, кроме пробела, табуляции, \n
\w - буква, цифра и _
\W - всё, кроме буквы, цифры и _
```

Кванторы могут быть жадными и не жадными. Примеры до этого были жадными, алгоритм работы прост - мы ищем символ на данной позиции, если совпадения нет, то переходим на следующую позицию. Когда символы для поиска заканчиваются, кол-во повторений уменьшается до нахождения следующего символа в регулярном выражении.

Для ленивого поиска после квантификатора указывается знак вопроса `?`. В таком случае мы будем повторять квантификатор наименьшее кол-во раз. После нахождения совпадения мы будем стараться искать следующий символ в выражении и только в случае, если мы его не найдём, кол-во повторений увеличится.

```
data = 'много текста (это нужный текст) ещё много текста (это тоже важно) ^^'

re.findall(r'\(.(+)\)', data)
>>> ['(это нужный текст) ещё много текста (это тоже важно)']

re.findall(r'\(.(+?)\)', data)
>>> ['(это нужный текст)', '(это тоже важно)']
```

Задания

Текст для поиска

```
dakak30@yahoo.com|false +77995231033
calan62@yandex.ru|false +77610387005
zanubij49@gmail.com|false +73308657933
N +7(916)9985670 Лев З. Зибefий
N 8-912-268-5440 Ян Ф. Балигук
N 8905148-3339 Лев У. Сонорин
8(913)448-51-90 Да Табушберг, А.К. 0.05
903-345-34-34 Да Нушешянц, Г.И. 0.24
903-34-334-34 Да Вефук, А.Г. 0.71
903-34-33434 Нет Мулигин, Ф.М. 0.08
Мефский, К.Д. 04/04/22 0.2
Мафусяк, Д.Р. 04/12/23 0.5
Вумук, С.С. 99/10/08 0.1
Вокезман, А.Е. 04/12/10 0.9
Не выполнено 1999-12-01 ramil_30@rambler.ru
Не выполнено 1999-12-01 ramil_30@rambler.ru
Не выполнено 1999-12-01 ramil_30@rambler.ru
Не выполнено 2001-09-12 busman96@yandex.ru
Не выполнено 2000-06-22 dmitrij47@yahoo.com
Выполнено 2002-05-17 vaceslav86@yandex.ru
```

Задание 1. Поиск email адресов

Необходимо найти все email адреса

Вывод должен получиться следующий:

```
dakak30@yahoo.com
calan62@yandex.ru
zanubij49@gmail.com
cuvij44@rambler.ru
ramil_30@rambler.ru
ramil_30@rambler.ru
ramil_30@rambler.ru
busman96@yandex.ru
dmitrij47@yahoo.com
vaceslav86@yandex.ru
```

Задание 2. Поиск номеров телефонов

По аналогии с первым заданием необходимо найти все номера мобильных телефонов.

Вывод должен получиться следующий:

```
+77995231033
+77610387005
+73308657933
```

```
+7(916)9985670  
8-912-268-5440  
8905148-3339  
8(913)448-51-90  
903-345-34-34  
903-34-334-34  
903-34-33434
```