

## Теоретический материал к занятию Введение в регулярные выражения 2 занятие.

### **re.split(шаблон, строка, максимальное кол-во разделений=0)**

Максимальное количество разделений, является опциональным и если мы его не укажем, то разбиение строки будет столько раз, сколько совпал шаблон с нашей строкой.

```
import re

sentence = 'Python. Я занимаюсь программированием на Python!'
result = re.split(r'Python', sentence)
print(result)
```

Результат

*['', '. Я занимаюсь программированием на ', '!']*

### **re.sub(шаблон, значение на которое меняем, строка)**

```
import re

sentence = 'Я занимаюсь программированием на PHP!'
result = re.sub(r'PHP', 'Python', sentence)
print(result)
```

Результат

*Я занимаюсь программированием на Python!*

Метод sub заменит все совпадающие с шаблоном последовательности символов в строке

```
import re

sentence = 'Java. Я занимаюсь программированием на Java!'
result = re.sub(r'Java', 'Python', sentence)
print(result)
```

Результат

*Python. Я занимаюсь программированием на Python!*

Можно собрать регулярное выражение в отдельный объект, который может быть использован для поиска. Это также избавит нас от переписывания одного и того же выражения.

```
import re

pattern = re.compile('Python')
result = pattern.findall('Python. Я занимаюсь программированием на Python!')
```

```
print(result)
result2 = pattern.findall('Я люблю язык программирования Python')
print(result2)
result3 = pattern.sub('Java', 'Я знаю язык программирования Python')
print(result3)
```

Основные специальные символы в регулярных выражениях:

.	Один любой символ, кроме новой строки \n.
?	0 или 1 вхождение шаблона слева
+	1 и более вхождений шаблона слева
*	0 и более вхождений шаблона слева
\w	Любая цифра или буква (\W – все, кроме буквы или цифры)
\d	Любая цифра [0-9] (\D – все, кроме цифры)
\s	Любой пробельный символ (\S – любой непробельный символ)
\b	Граница слова
[..]	Один из символов в скобках ([^..] – любой символ, кроме тех, что в скобках)
\	Экранирование специальных символов (\. означает точку или \+ – знак «плюс»)
^ и \$	Начало и конец строки соответственно
{n,m}	От n до m вхождений ({,m} – от 0 до m)
a b	Соответствует a или b
()	Группирует выражение и возвращает найденный текст
\t, \n, \r	Символ табуляции, новой строки и возврата каретки соответственно