

# Разбиения в регулярных выражениях

# Разбиение текста на отдельные слова

```
def get_tokens(text):
```

```
    tokens = text.split()
```

```
    # Долгая чистка всякого мусора
```

```
    return tokens
```

# Разбиение текста на отдельные слов

```
def get_tokens(text):  
    tokens = re.split('\W+', text)  
    return tokens
```

```
re.split(pattern, string, maxsplit=0, flags=0)
```

# maxsplit

```
text = 'приветливое приведение с причудами'
```

```
print(re.split('\W+', text))
```

```
print(re.split('\W+', text, maxsplit=1))
```

```
['приветливое', 'приведение', 'с', 'причудами']
```

```
['приветливое', 'приведение с причудами']
```

Найти все слова с ё. Что такое flags

# Антоним findall

```
text = 'Иван Родил Девчонку, Велел Тащить Пелёнку'
```

```
print(re.findall('\w+', text))
```

```
print(re.split('\W+', text))
```

```
['Иван', 'Родил', 'Девчонку', 'Велел', 'Тащить', 'Пелёнку']
```

```
['Иван', 'Родил', 'Девчонку', 'Велел', 'Тащить', 'Пелёнку']
```

## Группы сохраняются

```
expr = '3 + 2 * 5 / 33 - 2'
```

```
print(re.split('\s*[\+*\-/\]\s*', expr))
```

```
print(re.split('\s*([\+*\-/\])\s*', expr))
```

```
['3', '2', '5', '33', '2']
```

```
['3', '+', '2', '*', '5', '/', '33', '-', '2']
```



```
text = 'ЯШелВ7ЧасовВечераДомой'  
result = re.split('([0-9]+|[А-Я][а-я]*)', text)  
print(result)
```

<https://vk.cc/9eURWz>

# Задание

Написать программу, которая

1. Спросит путь к файлу, который надо считать
2. Найдёт там все слова с Ъ на конце. Распечатает их количество
3. Заменит все слова с Ъ на конце на слова без него
4. Сохранит полученный результат в файле с названием **<имя исходного файла>.proccessed.txt**
5. Сохранит в отдельный файл **<имя исходного файла>.replacements.txt** список уникальных слов с Ъ на конце с количеством таких слов в исходном файле в формате **<СЛОВО>|<КОЛИЧЕСТВО>**