

---

# rusclasp Documentation

*Выпуск 1*

gree-gorey

28 January 2017



<b>1</b>	<b>Документация</b>	<b>3</b>
1.1	Руководство пользователя . . . . .	3
1.2	Опции вывода . . . . .	4
<b>2</b>	<b>Устройство</b>	<b>5</b>
2.1	Предобработка . . . . .	5
2.2	Словари . . . . .	5



rusclasp — это python модуль, позволяющий разбивать предложения русского языка на предикации с восстановлением разрывов.

Содержание:



---

## Документация

---

## 1.1 Руководство пользователя

### 1.1.1 Установка из PyPI

Для установки используйте `pip`:

```
$ pip install rusclasp
```

#### Dependencies

Для работы `rusclasp` необходим `TreeTagger` (вместе с `Russian parameter file`) а также `treetaggerwrapper`, `treetaggerwrapper` может быть установлен следующей командой:

```
$ pip install treetaggerwrapper
```

### 1.1.2 Разбиение предикаций

Разбиение на предикации представляет собой разбиение текста на отрезки, содержащие один финитный предикат (возможно, нулевой) и его зависимые. При этом предикации, разорванные вложением, восстанавливаются.

Разбиение происходит следующим образом:

```
import rusclasp

s = rusclasp.Splitter()

sentence = u'Вы можете, введя свое предложение, проверить работу программы.'

result = s.split(sentence)
```

У метода `split` есть необязательный аргумент `mode`, значение которого по умолчанию `mode='json'`. В таком режиме он возвращает в переменную `result` словарь со следующей схемой:

```
{ 'text': "Текст, который вы разбиваете.",
  'entities': [
    ['T1', 'Span', [[0, 2]]],
    ['T2', 'Span', [[4, 8]]],
    ['T3', 'Span', [[10, 15]]]
```

```
    ],  
    'relations': [  
        ['R1', 'Split', [['LeftSpan', 'T1'], ['RightSpan', 'T3']]]  
    ]  
}
```

## 1.2 Опции вывода

### 1.2.1 Запись в файл

Запись

### 1.2.2 Запись разметки

Запись



## 2.1 Предобработка

### 2.1.1 POS-tagging

про POS-tagging

### 2.1.2 Разбиение на предложения

про разбиение

## 2.2 Словари

### 2.2.1 Союзы

про союзы

### 2.2.2 Вводные слова

про вводные слова

Исходный код - на [github](#). Если заметили ошибку, то пишите в баг-трекер.