NLP Basics: Learning how to use regular expressions

Using regular expressions in Python

Python's `re` package is the most commonly used regex resource. More details can be found here.

```
import re

re_test = 'This is a made up string to test 2 different regex methods'
re_test_messy = 'This      is a made up     string to test 2    different regex methods'
re_test_messy1 = 'This-is-a-made/up.string*to>>>>test----2""""""different~regex-methods'
```

Splitting a sentence into a list of words

```
re.split('\s', re_test)
```

```
['This',
 'is',
 'a',
 'made',
 'up',
 'string',
 'to',
 'test',
 '2',
 'different',
 'regex',
 'methods']
```

```
re.split('\s', re_test_messy)
```

```
['This',
 '',
 '',
 '',
 '',
```

```
      '',
      'is',
      'a',
      'made',
      'up',
      '',
      '',
      '',
      '',
      '',
      'string',
      'to',
      'test',
      '2',
      '',
      '',
      '',
      'different',
      'regex',
      'methods']


re.split('\s+', re_test_messy)

     ['This',
      'is',
      'a',
      'made',
      'up',
      'string',
      'to',
      'test',
      '2',
      'different',
      'regex',
      'methods']


re.split('\s+', re_test_messy1)

     ['This-is-a-made/up.string*to>>>>test----2"""""""different~regex-methods']


re.split('\W+', re_test_messy1)

     ['This',
      'is',
      'a',
```

```
       'made',
       'up',
       'string',
       'to',
       'test',
       '2',
       'different',
       'regex',
       'methods']


re.findall('\S+', re_test)

     ['This',
       'is',
       'a',
       'made',
       'up',
       'string',
       'to',
       'test',
       '2',
       'different',
       'regex',
       'methods']


re.findall('\S+', re_test_messy)

     ['This',
       'is',
       'a',
       'made',
       'up',
       'string',
       'to',
       'test',
       '2',
       'different',
       'regex',
       'methods']


re.findall('\S+', re_test_messy1)

     ['This-is-a-made/up.string*to>>>>test----2"""""""different~regex-methods']
```

```
re.findall('\w+', re_test_messy1)
```

```
['This',
 'is',
 'a',
 'made',
 'up',
 'string',
 'to',
 'test',
 '2',
 'different',
 'regex',
 'methods']
```

Replacing a specific string

```
pep8_test = 'I try to follow PEP8 guidelines'
pep7_test = 'I try to follow PEP7 guidelines'
peep8_test = 'I try to follow PEEP8 guidelines'
```

```
import re
```

```
re.findall('[a-z]+', pep8_test)
```

```
['try', 'to', 'follow', 'guidelines']
```

```
re.findall('[A-Z]+', pep8_test)
```

```
['I', 'PEP']
```

```
re.findall('[A-Z]+[0-9]+', peep8_test)
```

```
['PEEP8']
```

```
re.sub('[A-Z]+[0-9]+', 'PEP8 Python Styleguide', peep8_test)
```

```
'I try to follow PEP8 Python Styleguide guidelines'
```

Other examples of regex methods

- re.search()
- re.match()
- re.fullmatch()
- re.finditer()
- re.escape()