

## Groups and quantifiers in regular expressions

In python re, we cannot use a modifier on a capturing group. For example, if we would like to find sequences of 10 words in a row, with no punctuation, then we are looking for the pattern `(\w+ )10` times. The expression `(\w+ ){10}` (or any other quantifier) will not work here. We can only use quantifiers with non-capturing groups, denoted by `(?: )`. For example:

```
In: s='bb4b a1a2a3a4a5'
In: re.findall(r'(?:a\d)+',s)
Out: ['a1a2a3a4a5']
```

## Exercises

- Use regular expressions to write a function that takes a list as input and returns the length of the list.
- Write a regular expression that matches to sequences of two or more words starting with the letter 'a' in a piece of text.

## Replace

We can use the `replace()` function to replace substrings of a string with other substrings. There is an optional third argument *k*, which causes only the first *k* instances to be replaced.

```
In: s='aba'
In: s.replace('a','c')
Out: 'cbc'
In: s.replace('a','c',1)
Out: 'cba'
```

## Exercises

- Use only the `replace` function to turn the string 'aaaa' into the string 'abcd'.

## Reading URLs into strings

For one of the homework questions, we wish to import the data from a url into a string. We use the module `urllib` or `urllib2` for this, and the following code.

```
#python2
import urllib2
page=urllib2.urlopen(url).read()
page=str(page)
```

```
#python3
import urllib
page=urllib.request.urlopen(url).read()
page=str(page)
```