Pattern Matching

A RegEx, or Regular Expression, is a special sequence of characters that defines a search pattern for complex string-matching mechanism.

Python has a built-in module re that can be used to work with Regular Expressions.

Importing the "re" module:

```
import re
```

search() Function

The search() function searches the string for a match and returns a match object if there is a match.

If there is more than one match, only the first occurrence of the match will be returned.

match() Function:

Returns a match object if there is a match in the beginning of the string

Search the string to see if it starts with “Python":

```
import re
msg = "Python is an object-oriented language."
x = re.match("Python", msg)
if (x):
          print("Pattern is matched.")
else:
          print("No match")
```

```
Pattern is matched.
```

Search the string to see if it starts with "The" and ends with "Spain":

```
import re
#Check if the string starts with "The" and ends with "Spain":
txt = "The rain in Spain"
x = re.search("^The.*Spain$", txt)
if (x):
          print("YES! We have a match!")
else:
          print("No match")
```

```
YES! We have a match!
```

Search for the first white-space character in the string:

```
import re
txt = "Thexyz rain in Spain"
x = re.search("\s", txt)
```

```
print("The first white-space character is located in position:",
x.start())
```

```
The first white-space character is located in position: 6
```

```
#Make a search that returns no match:

import re
txt = "The rain in Spain"
x = re.search("Portugal", txt)
print(x)
```

```
None
```

sub() Function

The sub() function replaces the matches with the text of your choice:

```
import re
txt = "The rain in Spain"
x = re.sub("\s", "9", txt)
print(x)
```

```
The9rain9in9Spain
```

```
#Note: You can control the number of replacements by specifying the
count parameter:

#Replace the first 2 occurrences:

import re
txt = "The rain in Spain"
x = re.sub("\s", "9", txt, 2)
print(x)
```

```
The9rain9in Spain
```

findall() Function:

The findall() function returns a list containing all matches.

```
#Print a list of all matches:

import re
txt = "The rain in Spain"
x = re.findall("rai", txt)
print(x)
```

```
['rai']
```

split() Function:

The split() function returns a list where the string has been split at each match:

```
#Split at each white-space character:

import re
txt = "The rain in Spain"
x = re.split("\s", txt)
print(x)
```

```
['The', 'rain', 'in', 'Spain']
```

```
#Note: You can control the number of occurrences by specifying the
maxsplit parameter.
#Split the string only at the first occurrence:

import re
txt = "The rain in Spain"
x = re.split("\s", txt, 2)
print(x)
```

```
['The', 'rain', 'in Spain']
```