University of Tripoli
Faculty of Information Technology
Department of Software Engineering

ITSE305 مواضيع مختارة
Python Programming
S2025

Lecture (9): Python RegEx

Python RegEx

- A RegEx, or Regular Expression, is a sequence of characters that forms a search pattern.
- RegEx can be used to check if a string contains the specified search pattern.

RegEx Module

- Python has a built-in package called re, which can be used to work with Regular Expressions.
- Import the re module: `import re`
- When you have imported the re module, you can start using regular expressions

RegEx Functions

- There module offers a set of functions that allows us to search a string for a match:

Function	Description
findall	Returns a list containing all matches
search	Returns a Match object if there is a match anywhere in the string
split	Returns a list where the string has been split at each match
sub	Replaces one or many matches with a string

The findall() Function

- The findall() function returns a list containing all matches.
- The list contains the matches in the order they are found.
- If no matches are found, an empty list is returned

```
import re

def findIt(sub,txt):
 x = re.findall(sub, txt)
 if (x):
  print(f"Yes, there is at least one match of ({sub}) in the string ({txt})!")
 else:
  print(f"No match of ({sub}) in the string ({txt})")

findIt("ai","The rain in Spain")
findIt("Portugal", "The rain in Spain")
```

```
Yes, there is at least one match of (ai) in the string (The rain in Spain)!
No match of (Portugal) in the string (The rain in Spain)
```


The 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:
- If no matches are found, the value None is returned

```
import re

txt = "The rain in Spain"
x = re.search("\s", txt)

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

x = re.search("Portugal", txt)
print(x)
```

```
The first white-space character is located in position: 3
None
```

The split() Function

- The split() function returns a list where the string has been split at each match
- You can control the number of occurrences by specifying the maxsplit parameter

```
import re

#Split at each white-space character:
txt = "The rain in Spain"
x = re.split("\s", txt)
print(x)

#Split the string only at the first occurrence:
txt = "The rain in Spain"
x = re.split("\s", txt, 1)
print(x)
```

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


The sub() Function

- The sub() function replaces the matches with the text of your choice
- You can control the number of replacements by specifying the count parameter

```
import re

#Replace all white-space characters with --:

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

#Replace the first two occurrences of a white-space character with --:

x = re.sub("\s", "--", txt, 2)
print(x)
```

```
The--rain--in--Spain
The--rain--in Spain
```

Metacharacters

- Metacharacters are characters with a special meaning:

Character	Description	Example
[ ]	A set of characters	"[a-m]"
\	Signals a special sequence (can also be used to escape special characters)	"\d"
.	Any character (except newline character)	"he..o"
^	Starts with	"^hello"
$	Ends with	"planet$"
*	Zero or more occurrences	"he.*o"
+	One or more occurrences	"he.+o"
?	Zero or one occurrences	"he.?o"
{ }	Exactly the specified number of occurrences	"he.{2}o"
|	Either or	"falls|stays"
( )	Capture and group	

Metacharacters – Example (1)

```
import re

txt = "That will be 59 dollars"

#Find all digit characters:

x = re.findall("\d", txt)
print(x)
```

```
['5', '9']
```

Metacharacters – Example (2)

```
import re

txt = "hello planet"

#Check if the string starts with 'hello':

x = re.findall("^hello", txt)
if x:
  print("Yes, the string starts with 'hello'")
else:
  print("No match")
```

```
Yes, the string starts with 'hello'
```


Metacharacters – Example (3)

```
import re

txt = "The rain in Spain falls mainly in the plain!"

#Check if the string contains either "falls" or "stays":

x = re.findall("falls|stays", txt)

print(x)

if x:
  print("Yes, there is at least one match!")
else:
  print("No match")
```

```
['falls']
Yes, there is at least one match!
```

The END