RegEx & Pip

RegEx:

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.

Try yourself:

#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")

RegEx Functions

The re 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

RegEx & Pip

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

Code for RegEx Functions:

```
import re

#Return a list containing every occurrence of "ai":

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

```
['ai', 'ai']
```

```
import re

txt = "The 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: 3
```

```
import re

#Split the string at every white-space character:

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

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

```
import re

#Replace all white-space characters with the digit "9":

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

```
The9rain9in9Spain
```

RegEx & Pip

Special Sequence:

Character	Description
\A	Returns a match if the specified characters are at the beginning of the string
\b	Returns a match where the specified characters are at the beginning or at the end of a word (the "r" in the beginning is making sure that the string is being treated as a "raw string")
\B	Returns a match where the specified characters are present, but NOT at the beginning (or at the end) of a word (the "r" in the beginning is making sure that the string is being treated as a "raw string")
\d	Returns a match where the string contains digits (numbers from 0-9)
\D	Returns a match where the string DOES NOT contain digits
\s	Returns a match where the string contains a white space character
\S	Returns a match where the string DOES NOT contain a white space character
\w	Returns a match where the string contains any word characters (characters from a to Z, digits from 0-9, and the underscore _ character)
\W	Returns a match where the string DOES NOT contain any word characters
\Z	Returns a match if the specified characters are at the end of the string

Example:

```
import re

txt = "The rain in Spain"

#Check if "ain" is present, but NOT at the beginning of a word:

x = re.findall(r"\Bain", txt)

print(x)

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

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

Python PIP:

PIP is a package manager for Python packages, or module

RegEx & Pip

What is a Package?

A package contains all the files you need for a module.

Modules are Python code libraries you can include in your project.

Pip install CamelCase

```
Microsoft Windows [Version 10.0.19045.4651]
(c) Microsoft Corporation. All rights reserved.

C:\Users\DenilaRajendran>pip install CamelCase
Collecting CamelCase
  Downloading camelcase-0.2.tar.gz (1.3 kB)
  Preparing metadata (setup.py) ... done
Building wheels for collected packages: CamelCase
  Building wheel for CamelCase (setup.py) ... done
  Created wheel for CamelCase: filename=camelcase-0.2-py3-none-any.whl size=1779 sha256=93d2a7205524179882218fd284aea09c5e88fdbdf6c5bfefe3a3fd4f036cb1e1
  Stored in directory: c:\users\denilarajendran\appdata\local\pip\cache\wheels\a7\40\a3\900133dd6de3e10c219659fec4118138db05d778e519c0b2bc
Successfully built CamelCase
Installing collected packages: CamelCase
Successfully installed CamelCase-0.2

[notice] A new release of pip is available: 24.0 -> 24.2
[notice] To update, run: python.exe -m pip install --upgrade pip

C:\Users\DenilaRajendran>
```

```
import camelcase

c = camelcase.CamelCase()

txt = "lorem ipsum dolor sit amet"

print(c.hump(txt))

#This method capitalizes the first letter of each word.
```

```
Lorem Ipsum Dolor Sit Amet
```

Pip list

RegEx & Pip

```
C:\Users\DenilaRajendran>pip list
Package                                  Version
---------------------------------------- -------------
absl-py                                  2.1.0
anyio                                    4.3.0
argon2-cffi                              23.1.0
argon2-cffi-bindings                     21.2.0
arrow                                    1.3.0
asttokens                                2.4.1
astunparse                               1.6.3
async-lru                                2.0.4
attrs                                    23.2.0
azure-ai-documentintelligence            1.0.0b2
azure-ai-language-questionanswering      1.1.0
azure-ai-textanalytics                   5.3.0
azure-common                             1.1.28
azure-core                               1.30.2
Babel                                    2.15.0
beautifulsoup4                           4.12.3
bleach                                   6.1.0
camelcase                                0.2
certifi                                  2024.2.2
cffi                                     1.16.0
charset-normalizer                       3.3.2
colorama                                 0.4.6
comm                                     0.2.2
contourpy                                1.2.1
cycler                                   0.12.1
debugpy                                  1.8.1
decorator                                5.1.1
defusedxml                               0.7.1
executing                                2.0.1
fastjsonschema                           2.19.1
flatbuffers                              24.3.25
fonttools                                4.51.0
fqdn                                     1.5.1
gast                                     0.6.0
google-pasta                             0.2.0
```

RegEx & Pip

Thank you