

Python RegEx



Next >

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

RegEx in Python

When you have imported the re module, you can start using regular expressions:

```
Example

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

import re

txt = "The rain in Spain"

x = re.search("^The.*Spain$", txt)

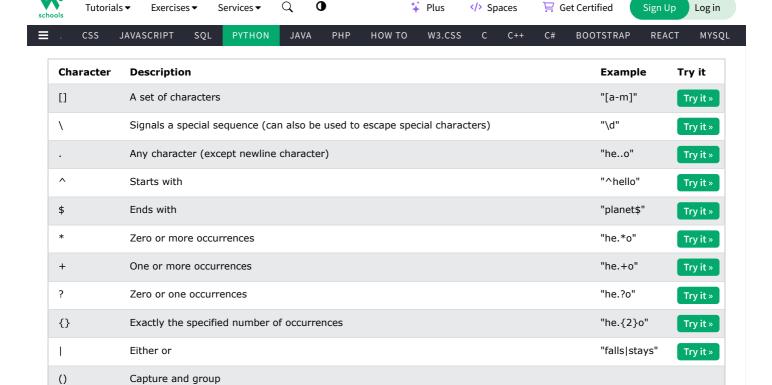
Try it Yourself »
```

RegEx Functions

The re module offers a set of functions that allows us to search a string for a match:

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

Services **▼**



¥ Plus

Get Certified

Special Sequences

Tutorials ▼

Exercises **▼**

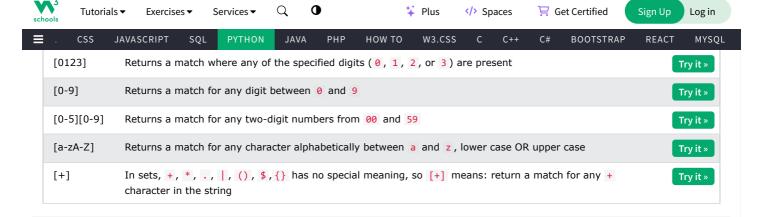
A special sequence is a \ followed by one of the characters in the list below, and has a special meaning:

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

Sets

A set is a set of characters inside a pair of square brackets [] with a special meaning:

Set	Description	Try it
[arn]	Returns a match where one of the specified characters (a , r , or n) is present	Try it »



The 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("ai", txt) print(x) Try it Yourself »

The list contains the matches in the order they are found.

If no matches are found, an empty list is returned:

Example

Return an empty list if no match was found:

```
import re

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

Try it Yourself »
```

The search() Function

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

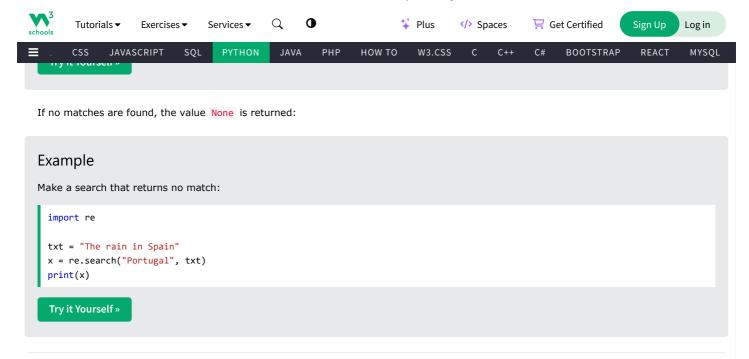
If there is more than one match, only the first occurrence of the match will be returned:

Example

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

```
import re

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



The split() Function

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

Example

Split at each white-space character:

```
import re

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

Try it Yourself »
```

You can control the number of occurrences by specifying the maxsplit parameter:

Example

Split the string only at the first occurrence:

```
import re

txt = "The rain in Spain"
x = re.split("\s", txt, 1)
print(x)

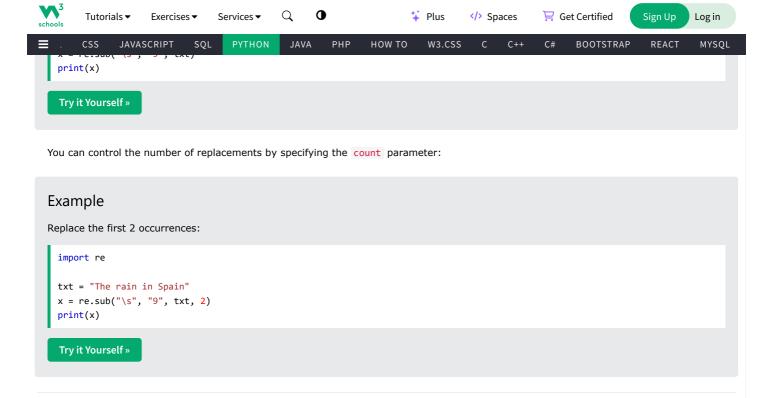
Try it Yourself »
```

The sub() Function

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

Example

Replace every white-space character with the number 9:



Match Object

A Match Object is an object containing information about the search and the result.

Note: If there is no match, the value None will be returned, instead of the Match Object.

Example

Do a search that will return a Match Object:

```
import re

txt = "The rain in Spain"
x = re.search("ai", txt)
print(x) #this will print an object

Try it Yourself »
```

The Match object has properties and methods used to retrieve information about the search, and the result:

```
.span() returns a tuple containing the start-, and end positions of the match.
```

- .string returns the string passed into the function
- .group() returns the part of the string where there was a match

Example

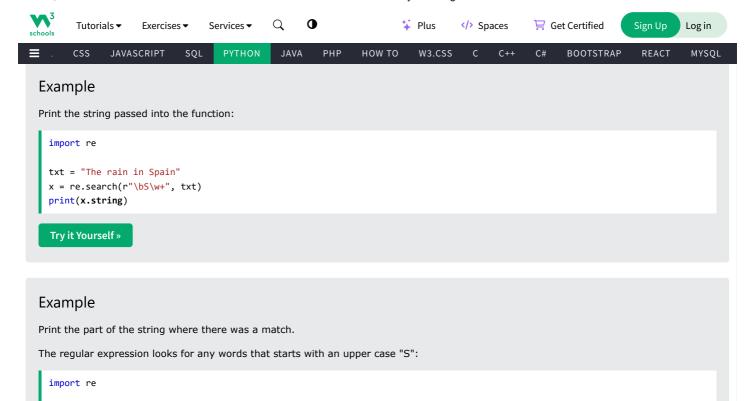
Print the position (start- and end-position) of the first match occurrence.

The regular expression looks for any words that starts with an upper case "S":

```
import re

txt = "The rain in Spain"

x = re.search(r"\bS\w+", txt)
print(x.span())
```



txt = "The rain in Spain"
x = re.search(r"\bS\w+", txt)
print(x.group())

Try it Yourself »

Note: If there is no match, the value None will be returned, instead of the Match Object.

```
Exercise

Consider the following code:

import re

txt = 'The rain in Spain'

x = re.findall('[a-c]', txt)

print(x)

What will be the printed result?

( 'a', 'a']

'The rin in Spin'

2

Submit Answer »
```

V3schools Pathfinder





COLOR PICKER





