REGEX IN PYTHON

Regular expression is a sequence of characters that forms a pattern which is mainly used to find or replace patterns in a string. These are supported by many languages such as python, java, R etc. Most common uses of regular expressions are:

- Finding patterns in a string or file.(Ex: find all the numbers present in a string)
- Replace a part of the string with another string.
- Search substring in string or file.
- Split string into substrings.
- Validate email format or passwords.

we need to **import** rejEx module **'re'** in our file.

Python RegEx Expressions Functions

SI NO	FUNCTION	DESCRIPTION
1	findall(pattern,string)	It gives all the occurrences of the pattern present in the string in the form of list.
2	search(pattern,string)	It gives that the pattern which is present at any position in the string.
3	split(pattern,string)	It splits the string on the given pattern.
4	sub(pattern,rep_substring,string)	It replaces one or more matching pattern in the string with the given substring.

Meta Characters

SI NO	META CHARACTER	DESCRIPTION
1	[]	**[a-e]** if you give like this it will give the letter from a to e. **[0-5]** if you give like this it will give o to 5 number in the given string.

2	\	It only get the numbers inside the given string and reject alphabets and special characters.
3	.	**“Testing”** from this word, if you find like **“Te...ng”**, then it will give the output like **“Testing”**
4	^	If the given string is start with the finding text, then it gives true.
5	$	If the given string is end with the finding text, then it gives true.
6	*	**“Testing can be done”** from this sentence, if you want to find from **“Tes....up to the last word done”,** instead of using **“.”** this, you can use **“.*”** like below. **“Tes.*done”.**
7	+	this one is same like the above but only one difference is **+** is one or more but ***** is zero or more occurrences
8	?	this one also same like above two but only difference is that it is zero or only one occurrences.
9	{}	with this braces, we can give the number of occurrences like below. for **“Testing”** word, **“Te.{4}g”**
10	|	this is for or condition
11	()	grouping of one or two regex pattern

SPECIAL SEQUENCE

SI NO	SEQUENCE	DESCRIPTION
1	\A	Gives true when the specified string in present in the beginning of the string.
2	\b	This gives true if the characters to the right are at the beginning of a word or the characters to the left are at the end of a word in the given string.
3	\B	opposite of \b

4	\d	This gives a match if the string contains a digit.
5	\D	This gives a match if the string contains only non digit characters.
6	\s	Returns a match where the string contains a white space character
7	\S	Returns a match where the string DOES NOT contain a white space character
8	\w	This gives a match if the string contains any character in a-z, A-Z, 0-9 and underscore(_).
9	\W	This gives a match if the string contains characters other than a-z, A-Z, 0-9 and underscore(_).
10	\Z	This gives a match if the characters to the left of \Z are present at the end of the string.

SETS

SI NO	SETS	DESCRIPTION
1	[abcd]	Gives a match if the string contains a,b,c or d.
2	[a-z]	Gives a match if the string contains any character from a to z.
3	[A-Z]	Gives a match if the string contains any character from A to Z.
4	[0-9]	Gives a match if string contains digits from 0 to 9
5	[a-zA-Z0-9]	Gives a match if any of the above conditions holds true.
6	[^a-zA-Z]	Gives a match if the string doesn’t contain any alphabet.

FINDALL FUNCTION

```
import re
```

```
'
```

```
# findall() function returns a list which contains all matches
data = "my institute name is softlogic"

result = re.findall("institute", data)
print(result)

result = re.findall("soft tech", data)
print(result)

result = re.findall("^my", data)
print(result)

result = re.findall("logic$", data)
print(result)
```

SEARCH FUNCTION

```
import re

# search() function gives only the first occurrence of the string
data = "my institute name institute is softlogic"

result = re.search("institute", data)
print("start", result.start())

result = re.search("institute", data)
print("end", result.end())

result = re.search("^my", data)
print("span", result.span())

result = re.search("logic$", data)
print("group", result.group())
```

SPLIT FUNCTION

```
import re

# split function is used for splitting based on the pattern

data = "my institute name is softlogic"

# split with white space characters
```

```
# split with white space characters
result = re.split("\s", data)
print(result)

# split with 'name'
result = re.split("name", data)
print(result)
```

SUB FUNCTION

This function replaces a pattern with the given substring in a given string.

```
import re

# split function is used for splitting based on the pattern

data = "my institute name is softlogic"


# sub function
data = "my institute name is softlogic"

result = re.sub("softlogic", "SLA", data)
print(result)
```

MATCH FUNCTION

```
import re

pattern = r"\d+"
text = "12345"

match = re.match(pattern, text)
if match:
    print("Match found!")
else:
    print("No match found")
```