Python Programming

RGM College of Engineering & Technology (Autonomous)

Department of Computer Science & Engineering

Academic Year : 2020-2021

REGULAR EXPRESSIONS - 2

Guido Van Rossum

Learning Mantra

If you really strong in the basics, then remaining things will become so easy.

Agenda

1. Character classes
2. Pre defined Character classes
3. Quantifiers

3. Character classes

We can use character classes to search a group of characters.

1. [abc] ➔ Either a or b or c
2. [^abc] ➔ Except a and b and c
3. [a-z] ➔ Any Lower case alphabet symbol
4. [A-Z] ➔ Any upper case alphabet symbol
5. [a-zA-Z] ➔ Any alphabet symbol
6. [0-9] ➔ Any digit from 0 to 9
7. [a-zA-Z0-9] ➔ Any alphanumeric character
8. [^a-zA-Z0-9] ➔ Except alphanumeric characters(Special Characters)

Eg 1:

```
import re
matcher=re.finditer("[abc]","a7b@k9z")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
0 ...... a
2 ...... b
```

Eg 2:

```
import re
matcher=re.finditer("[^abc]","a7b@k9z")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
1 ...... 7
3 ...... @
4 ...... k
5 ...... 9
6 ...... z
```

Eg 3:

```
import re
matcher=re.finditer("[a-z]","a7b@k9z")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
0 ..... a
2 ..... b
4 ..... k
6 ..... z
```

Eg 4:

```
import re
matcher=re.finditer("[0-9]","a7b@k9z")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
1 ..... 7
5 ..... 9
```

Eg 6:

```
import re
matcher=re.finditer("[A-Z]","a7b@k9z")
for match in matcher:
    print(match.start(),"......",match.group())      # Nothing
```

Eg 7:

```
import re
matcher=re.finditer("[a-zA-Z]","a7b@k9z")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
0 ..... a
2 ..... b
4 ..... k
6 ..... z
```

Eg 8:

```
import re
matcher=re.finditer("[a-zA-Z0-9]","a7b@k9z")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
0 ...... a
1 ...... 7
2 ...... b
4 ...... k
5 ...... 9
6 ...... z
```

Eg 9:

```
import re
matcher=re.finditer("[^a-zA-Z0-9]","a7b@k9z")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
3 ...... @
```

Eg 10:

```
import re
matcher=re.finditer("[abc]","abcabc")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
0 ...... a
1 ...... b
2 ...... c
3 ...... a
4 ...... b
5 ...... c
```

4. Pre defined Character classes

- \s ➔ Space character
- \S ➔ Any character except space character
- \d ➔ Any digit from 0 to 9
- \D ➔ Any character except digit
- \w ➔ Any word character [a-zA-Z0-9]
- \W ➔ Any character except word character (only Special Characters includes)
- . ➔ Any character including special characters

Eg 1:

```
import re
matcher=re.finditer("\s","a7b k@9z")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
3 ..... 
```

Eg 2:

```
import re
matcher=re.finditer("\S","a7b k@9z")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
0 ..... a
1 ..... 7
2 ..... b
4 ..... k
5 ..... @
6 ..... 9
7 ..... z
```

Eg 3:

```
import re
matcher=re.finditer("\d","a7b k@9z")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
1 ...... 7
6 ...... 9
```

Eg 4:

```
import re
matcher=re.finditer("\D","a7b k@9z")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
0 ...... a
2 ...... b
3 ......
4 ...... k
5 ...... @
7 ...... z
```

Eg 5:

```
import re
matcher=re.finditer("\w","a7b k@9z")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
0 ...... a
1 ...... 7
2 ...... b
4 ...... k
6 ...... 9
7 ...... z
```

Eg 6:

```
import re
matcher=re.finditer("\W","a7b k@9z")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
3 ......
5 ...... @
```

Eg 7:

```
import re
matcher=re.finditer(".","a7b k@9z")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
0 ...... a
1 ...... 7
2 ...... b
3 ......
4 ...... k
5 ...... @
6 ...... 9
7 ...... z
```

5. Quantifiers

→ We can use quantifiers to specify the number of occurrences to match.

- a ➔ Exactly one 'a'
- a+ ➔ At least one 'a'
- a* ➔Any number of a's including zero number
- a? ➔ At most one 'a', i.e., either zero number or one number
- a{m} ➔ Exactly ‘m’ number of a's
- a{m,n} ➔ Minimum ‘m’ number of a's and Maximum ‘n’ number of a's

Eg 1:

```
import re
matcher=re.finditer("a","abaabaaab")       # Exactly one 'a'
for match in matcher:
    print(match.start(),"......",match.group())
```

```
0 ...... a
2 ...... a
3 ...... a
5 ...... a
6 ...... a
7 ...... a
```

Eg 2:

```
import re
matcher=re.finditer("a+","abaabaaab") # Atleast one 'a'
for match in matcher:
    print(match.start(),"......",match.group())
```

```
0 ...... a
2 ...... aa
5 ...... aaa
```

Eg 3:

```
import re
matcher=re.finditer("a*","abaabaaab") # Any no.of 'a's including zero
for match in matcher:
    print(match.start(),"......",match.group())
```

```
0 ...... a
1 ......
2 ...... aa
4 ......
5 ...... aaa
8 ......
9 ......
```

Eg 4:

```
import re
matcher=re.finditer("a?","abaabaaab") # Atmost one 'a',
for match in matcher:
    print(match.start(),"......",match.group())
```

```
0 ...... a
1 ......
2 ...... a
3 ...... a
4 ......
5 ...... a
6 ...... a
7 ...... a
8 ......
9 ......
```

Eg 5:

```
import re
matcher=re.finditer("a{3}","abaabaaab") # Exactly '3' number of 'a's
for match in matcher:
    print(match.start(),"......",match.group())
```

```
5 ...... aaa
```

Eg 6:

```
import re
matcher=re.finditer("a{2,4}","abaabaaab") # Minimum '2' 'a's and Maximum 4 ‘a’s
for match in matcher:
    print(match.start(),"......",match.group())
```

```
2 ...... aa
5 ...... aaa
```

Eg 7:

```
import re
matcher=re.finditer("a{2,2}","abaabaaab") # Minimum '2' 'a's and Maximum ‘2’ ‘a’s
for match in matcher:
    print(match.start(),"......",match.group())
```

```
2 ...... aa
5 ...... aa
```

Eg 8:

```
import re
matcher=re.finditer("a{1,4}","abaabaaab")
for match in matcher:
    print(match.start(),"......",match.group())
```

```
0 ...... a
2 ...... aa
5 ...... aaa
```

Eg 9:

```
import re
matcher=re.finditer("a{2}a*","abaabaaab") # Exactly'2' 'a's followed by any no. of ‘a’ s
for match in matcher:
    print(match.start(),"......",match.group())
```

```
2 ...... aa
5 ...... aaa
```

Eg 10:

```
import re
matcher=re.finditer("[^a]","abaabaaab")        # Except 'a'
for match in matcher:
    print(match.start(),"......",match.group())
```

```
1 ...... b
4 ...... b
8 ...... b
```

Note:

^x ➔ It will check whether target string starts with **x** or not.

x$ ➔ It will check whether target string ends with **x** or not.

Eg 1:

```
import re
matcher=re.finditer("^a","abaabaaab") # Whether the given string starts with ‘a’ or not
for match in matcher:
    print(match.start(),"......",match.group())
```

```
0 ...... a
```

Eg 2:

```
import re
matcher=re.finditer("a$","abaabaaab") # Whether the given string ends with ‘a” or not.
for match in matcher:
    print(match.start(),"……",match.group())          # Nothing
```

Eg 3:

```
import re
matcher=re.finditer("b$","abaabaaab") # Whether the given string ends with ‘b’ or not
for match in matcher:
    print(match.start(),"……",match.group())
```

```
8 …… b
```

Any question?

If you try to practice programs yourself, then you will learn many things automatically

Spend few minutes and then enjoy the study

Thank You