Regular expression practice questions

Name – subham

Batch – DS2401

Question 1- Write a RegEx pattern in python program to check that a string contains only a certain set of characters (in this case a-z, A-Z and 0-9).

Answer : we can use findall and search functions for this. patterns can be

re.findall(r'a-zA-Z0-9+', string) this pattern means first letter can be a-z or A-Z or 0-9 and + means it will look for same pattern to successive letters and if pattern not found (like wide space) it will stop and the show the matched pattern in output and string is the set of string character where it will search the pattern or

re.search(r'a-zA-Z0-9+', string)

```
In [19]: string='The Hyundai Creta also known as Hyundai ix25 in China is a subcompact crossover SUV'
         re.findall(r'[a-zA-Z0-9]+',string)

Out[19]: ['The',
          'Hyundai',
          'Creta',
          'also',
          'known',
          'as',
          'Hyundai',
          'ix25',
          'in',
          'China',
          'is',
          'a',
          'subcompact',
          'crossover',
          'SUV']
```

Question 2- Write a RegEx pattern that matches a string that has an a followed by zero or more b's

Answer : we can use findall and search method

Re.findall(r ' ab* ', string) or re.search(r ' ab* ', string)

Pattern = r'ab*' . r = raw string , pattern starts with a followed by b, * means it matches 0 or more repetitions of b as * is in front of b. if this pattern is in between a word it will also be matched as ^ symbol is not used which tells that a should be at the start(^ab*)

```
In [20]: string='a ab abbb abbc abbbbb alll'
         re.findall(r'ab*',string)

Out[20]: ['a', 'ab', 'abbb', 'abb', 'abbbbb', 'a']
```

Question 3- Write a RegEx pattern that matches a string that has an a followed by one or more b's

Answer : re.findall(r ' ab+ ' , string) or re.search (r ' ab+ ' , string)

Pattern = r'ab+' . r = raw string , pattern starts with a followed by b, + means it matches 1 or more repetitions of b as b is followed by + . if this pattern is in between a word it will also be matched as ^ symbol is not used which tells that a should be at the start(^ab+)

```
In [21]: string='a ab abbb abbc abbbbb alll'
         re.findall(r'ab+',string)

Out[21]: ['ab', 'abbb', 'abb', 'abbbbb']
```

Question 4- Write a RegEx pattern that matches a string that has an a followed by zero or one 'b'.

Answer : re.findall(r ' ab? ' , string) or re.search(r ' ab? ' , string)

Pattern = r'ab?' . r = raw string , pattern starts with a followed by b, ? means it matches 0 or 1 repetitions of b as b is followed by ? . if this pattern is in between a word it will also be matched as ^ symbol is not used which tells that a should be at the start(^ab?)

```
In [22]: string='a ab abbb abbc abbbbb alll'
         re.findall(r'ab?',string)

Out[22]: ['a', 'ab', 'ab', 'ab', 'ab', 'a']
```

Question 5- Write a RegEx pattern in python program that matches a string that has an a followed by three 'b'.

Answer :

import re

string= ' a ab abbb abbc abbbbb alll '

re.findall(r ' abbb ' , string)

Pattern = r'abbb' or r'ab{3}' . r = raw string , pattern starts with a followed by three b or b{3}. if this pattern is in between a word it will also be matched as ^ symbol is not used

```
In [25]: string= 'a ab abbb abbc abbbbb alll'
         re.findall(r'ab{3}',string)

Out[25]: ['abbb', 'abbb']
```

```
In [40]: string= 'a ab abbb abbc abbbbb alll'
         re.findall(r'abbb',string)

Out[40]: ['abbb', 'abbb']
```

Question 6- Write a RegEx pattern in python program that matches a string that has an a followed by two to three 'b'.

Answer :

string= ' a ab abbb abbc abbbbb alll '

re.findall(r' abbb? ' , string)

Pattern = r'abbb?' or r'ab{2,3}' . r = raw string , pattern starts with a followed by three b and ? or b{3}. if this pattern is in between a word it will also be matched as ^ symbol is not used at the start

```
In [24]: string= 'a ab abbb abbc abbbbb alll'
         re.findall(r'ab{2,3}',string)

Out[24]: ['abbb', 'abb', 'abbb']
```

```
In [41]: string= 'a ab abbb abbc abbbbb alll'
         re.findall(r'abbb?',string)

Out[41]: ['abbb', 'abb', 'abbb']
```

Question 7- Write a Python program that matches a string that has an 'a' followed by anything, ending in 'b'.

Answer : pattern = r'a.*b$' . a is followed by dot(.) which is followed by asterisk(*)which is followed by b$. dot matches any character except newline character . dollar($) matches b at the end of string. Asterisk(*) matches 0 or more repetitions of dot(.) 1st example starts with a and end with b so it is shown in output. 2nd example a is between a string and ends with b so it is shown in output. 3rd example its starts with a but doesn't end with b so not shown in output.

```
In [43]: string= 'adpejeb'
         re.findall(r'a.*b$',string)

Out[43]: ['adpejeb']
```

```
In [42]: string='jshdakb'
         re.findall(r'a.*b$',string)

Out[42]: ['akb']
```

```
In [40]: string='asbjhd'
         re.findall(r'a.*b$',string)

Out[40]: []
```

Question 8- Write a RegEx pattern in python program that matches a word at the beginning of a string.

Answer :

Pattern = r'^\w+' . \w matches where string is a word. ^ matches \w at the start. + matches \w until white space. Output comes out to be 'learning' .

In the 2nd example there is white space and in 3rd example there is special charcter at the beginning so output doesn't show anything

```
In [67]: string = 'learning must be a life-long process'
         re.findall(r'^\w+',string)

Out[67]: ['learning']
```

```
In [68]: string = ' learning must be a life-long process'
         re.findall(r'^\w+',string)

Out[68]: []
```

```
In [3]: string = '*learning must be a life-long process'
        re.findall(r'^\w+',string)

Out[3]: []
```

Question 9- Write a RegEx pattern in python program that matches a word at the end of a string.

Answer :

Pattern = r'\w+$' . \w matches where string is a word. $ matches \w at the end. + matches \w until white space. Output comes out to be 'process' .

In the 2nd example there is white space and in 3rd example there is special charcter at the end so output doesn't show anything

```
In [7]: string = 'learning must be a life-long process'
        re.findall(r'\w+$',string)

Out[7]: ['process']
```

```
In [3]: string = 'learning must be a life-long process '
        re.findall(r'\w+$',string)

Out[3]: []
```

```
In [4]: string = 'learning must be a life-long process%^&*;;'
        re.findall(r'\w+$',string)

Out[4]: []
```

Question 10- Write a RegEx pattern in python program to find all words that are 4 digits long in a string.

Answer :

Pattern = r'(\d{4})' . \d matches string that is digit . {4} specifies it should be 4 digits

```
In [13]: sample_text='01 0132 231875 1458 301 2725.'
         re.findall(r'(\d{4,4})',sample_text)

Out[13]: ['0132', '2318', '1458', '2725']
```