Regular Expressions

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

In [1]:
```
import regex as re
```

In [2]:
```
def is_allowed_specific_char(string):
    charRe = re.compile(r'[^a-zA-Z0-9.]')
    string = charRe.search(string)
    return not bool(string)
```

In [3]:
```
print(is_allowed_specific_char("ABCDEFabcdef123450"))
```

```
True
```

In [4]:
```
print(is_allowed_specific_char("*&%@#!}{"))
```

```
False
```

Question 2- Create a function in python that matches a string that has an a followed by zero or more b's

In [5]:
```
import regex as re
```

In [10]:
```
import re
def text_match(text):
        patterns = '^a(b*)$'
        if re.search(patterns,  text):
                return 'Found a match!'
        else:
                return('Not matched!')
```

In [11]:
```
print(text_match("ac"))
```

```
Not matched!
```

In [12]:
```
print(text_match("abc"))
```

```
Not matched!
```

In [13]:
```
print(text_match("a"))
```

```
Found a match!
```

In [14]:
```
print(text_match("ab"))
```

```
Found a match!
```

In [15]:
```
print(text_match("abb"))
```

```
Found a match!
```

Question 3- Create a function in python that matches a string that has an a followed by one or more b's

In [16]:
```
import re
def text_match(text):
        patterns = '^a(b*)$'
        if re.search(patterns,  text):
```

```
                return 'Found a match!'
        else:
                return('Not matched!')
```

In [17]:
```
print(text_match("ac"))
```

```
Not matched!
```

In [18]:
```
print(text_match("abc"))
```

```
Not matched!
```

In [19]:
```
print(text_match("a"))
```

```
Found a match!
```

In [20]:
```
print(text_match("ab"))
```

```
Found a match!
```

In [21]:
```
print(text_match("abb"))
```

```
Found a match!
```

Question 4- Create a function in Python and use RegEx that matches a string that has an a followed by zero or one 'b'.

In [22]:
```
import re

def match_string(string):
    pattern = r'a(b?)'
    match = re.search(pattern, string)
    if match:
        return True
    else:
        return False
```

In [23]:
```
print(text_match("ac"))
```

```
Not matched!
```

In [24]:
```
print(text_match("abc"))
```

```
Not matched!
```

In [25]:
```
print(text_match("a"))
```

```
Found a match!
```

In [26]:
```
print(text_match("ab"))
```

```
Found a match!
```

In [27]:
```
print(text_match("abb"))
```

```
Found a match!
```

Question 5- Write a Python program that matches a string that has an a followed by three 'b'.

In [28]:
```
import re
def text_match(text):
        patterns = '^a(b*)$'
        if re.search(patterns,  text):
```

```
                return 'Found a match!'
        else:
                return('Not matched!')
```

In [29]:
```
print(text_match("ac"))
```

```
Not matched!
```

In [30]:
```
print(text_match("abc"))
```

```
Not matched!
```

In [31]:
```
print(text_match("a"))
```

```
Found a match!
```

In [32]:
```
print(text_match("ab"))
```

```
Found a match!
```

In [33]:
```
print(text_match("abb"))
```

```
Found a match!
```

Question 6- Write a regular expression in Python to split a string into uppercase letters. Sample text: “ImportanceOfRegularExpressionsInPython” Output: [‘Importance’, ‘Of’, ‘Regular’, ‘Expression’, ‘In’, ‘Python’]

In [34]:
```
import regex as re
```

In [35]:
```
text = "ImportanceOfRegularExpressionsInPython"
result = re.findall('[A-Z][^A-Z]*', text)

print(result)
```

```
['Importance', 'Of', 'Regular', 'Expressions', 'In', 'Python']
```

Question 7- Write a Python program that matches a string that has an a followed by two to three 'b'.

In [36]:
```
import re
def text_match(text):
        patterns = 'ab{2,3}'
        if re.search(patterns,  text):
                return 'Found a match!'
        else:
                return('Not matched!')
```

In [37]:
```
print(text_match("ab"))
```

```
Not matched!
```

In [38]:
```
print(text_match("aabbbbbc"))
```

```
Found a match!
```

Question 8- Write a Python program to find sequences of lowercase letters joined with a underscore.

In [39]:
```
import re
def text_match(text):
        patterns = '^[a-z]+_[a-z]+$'
```

```
        if re.search(patterns,  text):
                return 'Found a match!'
        else:
                return('Not matched!')
```

In [40]:
```
print(text_match("aab_cbbbc"))
```

```
Found a match!
```

In [41]:
```
print(text_match("aab_Abbbc"))
```

```
Not matched!
```

In [42]:
```
print(text_match("Aaab_abbbc"))
```

```
Not matched!
```

Question 9- Write a Python program that matches a string that has an 'a' followed by anything, ending in 'b'.

In [43]:
```
import re
def text_match(text):
        patterns = 'a.*?b$'
        if re.search(patterns,  text):
                return 'Found a match!'
        else:
                return('Not matched!')
```

In [44]:
```
print(text_match("aabAbbbc"))
```

```
Not matched!
```

In [45]:
```
print(text_match("aabbbbd"))
```

```
Not matched!
```

In [46]:
```
print(text_match("accddbbjjjb"))
```

```
Found a match!
```

Question 10- Write a Python program that matches a word at the beginning of a string.

In [47]:
```
import re
def text_match(text):
        patterns = '^\w+'
        if re.search(patterns,  text):
                return 'Found a match!'
        else:
                return('Not matched!')
```

In [48]:
```
print(text_match("Uttrakhand is know by dev bhoomi"))
```

```
Found a match!
```

In [49]:
```
print(text_match(" Uttrakhand is know by dev bhoomi"))
```

```
Not matched!
```

Question 11- Write a Python program to match a string that contains only upper and lowercase letters, numbers, and underscores.

In [50]:
```
import re
def text_match(text):
```

```
        patterns = '^[a-zA-Z0-9_]*$'
        if re.search(patterns,  text):
                return 'Found a match!'
        else:
                return('Not matched!')
```

In [51]:
```
print(text_match("Uttrakhand is know by dev bhoomi."))
```

```
Not matched!
```

In [52]:
```
print(text_match("Python_Exercises_1"))
```

```
Found a match!
```

Question 12- Write a Python program where a string will start with a specific number.

In [61]:
```
import re

def match_num(string):
    text = re.compile(r"^1")
    if text.match(string):

        return 'Found a match!'
    else:

        return('Not matched!')
```

In [62]:
```
print(match_num('1-2345861'))
```

```
Found a match!
```

In [63]:
```
print(match_num('2-2345861'))
```

```
Not matched!
```

Question 13- Write a Python program to remove leading zeros from an IP address

In [64]:
```
import regex as re
```

In [65]:
```
def remove_zeros_from_ip(ip_add):
    new_ip_add = ".".join([str(int(i)) for i in ip_add.split(".")])
    return new_ip_add
```

In [66]:
```
print(remove_zeros_from_ip("255.024.01.01"))
```

```
255.24.1.1
```

In [67]:
```
print(remove_zeros_from_ip("127.0.0.01"))
```

```
127.0.0.1
```

Question 14- Write a regular expression in python to match a date string in the form of Month name followed by day number and year stored in a text file. Sample text : ' On August 15th 1947 that India was declared independent from British colonialism, and the reins of control were handed over to the leaders of the Country'. Output- August 15th 1947 Hint- Use re.match() method here

In [68]:
```
import re

text = "on august 15th 1947 that india was declared independent from british colon
match = re.match(r"([a-zA-Z]+) (\d+)(st|nd|rd|th) (\d{4})", text)
```

```
if match:
    print(match.group())
```

Question 15- Write a Python program to search some literals strings in a string. Go to the editor Sample text : 'The quick brown fox jumps over the lazy dog.' Searched words : 'fox', 'dog', 'horse'

In [73]:
```
import re
patterns = [ 'fox', 'dog', 'horse' ]
text = 'The quick brown fox jumps over the lazy dog.'
for pattern in patterns:
    print('Searching for "%s" in "%s" ->' % (pattern, text),)
    if re.search(pattern,  text):
        print('Matched!')
    else:
        print('Not Matched!')
```

```
Searching for "fox" in "The quick brown fox jumps over the lazy dog." ->
Matched!
Searching for "dog" in "The quick brown fox jumps over the lazy dog." ->
Matched!
Searching for "horse" in "The quick brown fox jumps over the lazy dog." ->
Not Matched!
```

Question 16- Write a Python program to search a literals string in a string and also find the location within the original string where the pattern occurs Sample text : 'The quick brown fox jumps over the lazy dog.' Searched words : 'fox'

In [74]:
```
import re
pattern = 'fox'
text = 'The quick brown fox jumps over the lazy dog.'
match = re.search(pattern, text)
s = match.start()
e = match.end()
print('Found "%s" in "%s" from %d to %d ' % \
    (match.re.pattern, match.string, s, e))
```

```
Found "fox" in "The quick brown fox jumps over the lazy dog." from 16 to 19
```

Question 17- Write a Python program to find the substrings within a string. Sample text : 'Python exercises, PHP exercises, C# exercises' Pattern : 'exercises'.

In [75]:
```
import re
text = 'Python exercises, PHP exercises, C# exercises'
pattern = 'exercises'
for match in re.findall(pattern, text):
    print('Found "%s"' % match)
```

```
Found "exercises"
Found "exercises"
Found "exercises"
```

Question 18- Write a Python program to find the occurrence and position of the substrings within a string.

In [76]:
```
import re

text = "This is a sample string"
pattern = "is"
```

```
for match in re.finditer(pattern, text):
    s = match.start()
    e = match.end()
    print(f"Found \"{text[s:e]}\" at {s}:{e}")
```

```
Found "is" at 2:4
Found "is" at 5:7
```

Question 19- Write a Python program to convert a date of yyyy-mm-dd format to dd-mm-yyyy format.

yyyy-mm-dd stands for year-month-day .

In [78]:
```
import re

def change_date_format(dt):
    return re.sub(r'(\\d{4})-(\\d{1,2})-(\\d{1,2})', '\\\\3-\\\\2-\\\\1', dt)

dt = "2023-07-13"
print("Original date in YYY-MM-DD Format: ", dt)
print("New date in DD-MM-YYYY Format: ", change_date_format(dt))
```

```
Original date in YYY-MM-DD Format:  2023-07-13
New date in DD-MM-YYYY Format:  2023-07-13
```

Question 20- Write a Python program to find all words starting with 'a' or 'e' in a given string.

In [79]:
```
import re

text = "Happiness is defined by different people in different ways. When we feel p

#find all the words starting with 'a' or 'e'
list = re.findall("[ae]\\w+", text)

#print result
print(list)
```

```
['appiness', 'efined', 'erent', 'eople', 'erent', 'ays', 'en', 'eel', 'emotions',
'end', 'eel', 'appy', 'at', 'at', 'appiness', 'all', 'about', 'appiness', 'also',
'egarded', 'as', 'ental', 'ate', 'erson', 'an', 'anner']
```

Question 21- Write a Python program to separate and print the numbers and their position of a given string.

In [4]:
```
import re

def separate_numbers(string):
    for match in re.finditer(r'\d+', string):
        print(f"Number {match.group()} found at position {match.start()}")

separate_numbers("Hello 987 Python 879")
```

```
Number 987 found at position 6
Number 879 found at position 17
```

In [5]:
```
import re
# Input.
text = "The population of India is currently around 1.4 billion. This makes India t

for m in re.finditer("\d+", text):
    print(m.group(0))
    print("Index position:", m.start())
```

```
1
Index position: 44
4
Index position: 46
1
Index position: 227
5
Index position: 229
2050
Index position: 242
```

Question 22- Write a regular expression in python program to extract maximum numeric value from a string

In [7]:
```python
import re

def extract_max_numeric_value(string):
    pattern = r'\d+'
    numbers = re.findall(pattern, string)
    max_num = max(map(int, numbers))
    return max_num
```

Question 23- Write a Regex in Python to put spaces between words starting with capital letters

In [25]:
```python
import re

def add_spaces(s):
    return re.sub(r'(?<!^)(?=[A-Z])', ' ', s)

s = "MyNameIsManishaSingh"
print(add_spaces(s))
```

```
My Name Is Manisha Singh
```

Question 24- Python regex to find sequences of one upper case letter followed by lower case letters

In [26]:
```python
import re

string = "This is a sample string with AaBbCcDdEeFfGgHhIiJjKkLlMmNnOoPpQqRrSsTtUuV
pattern = r'[A-Z][a-z]+'

result = re.findall(pattern, string)
print(result)
```

```
['This', 'Aa', 'Bb', 'Cc', 'Dd', 'Ee', 'Ff', 'Gg', 'Hh', 'Ii', 'Jj', 'Kk', 'Ll',
'Mm', 'Nn', 'Oo', 'Pp', 'Qq', 'Rr', 'Ss', 'Tt', 'Uu', 'Vv', 'Ww', 'Xx', 'Yy', 'Z
z']
```

Question 25- Write a Python program to remove duplicate words from Sentence using Regular Expression

In [28]:
```python
import re

def remove_duplicates(text):
    return re.sub(r'\b(\w+)(?:\W+\1\b)+', r'\1', text, flags=re.IGNORECASE)

sentence = "My My Name Name Is Is Manisha"
print(remove_duplicates(sentence))
```

```
My Name Is Manisha
```

Question 26- Write a python program using RegEx to accept string ending with alphanumeric character.

In [29]:
```
import re

def check_alphanumeric(string):
    regex = re.compile(r'^.*[a-zA-Z0-9]$')
    if regex.match(string):
        print("String ends with an alphanumeric character")
    else:
        print("String does not end with an alphanumeric character")
```

In [30]:
```
check_alphanumeric("Hello India 1234")
```

```
String ends with an alphanumeric character
```

In [31]:
```
check_alphanumeric("Hello Python")
```

```
String ends with an alphanumeric character
```

In [32]:
```
check_alphanumeric("Hello Python 9876")
```

```
String ends with an alphanumeric character
```

Question 27-Write a python program using RegEx to extract the hashtags.

In [34]:
```
import re

def extract_hashtags(text):
    hashtags = re.findall(r'\#\w+', text)
    return hashtags

text = "This is a #Dance string with #competition"
hashtags = extract_hashtags(text)
print(hashtags)
```

```
['#Dance', '#competition']
```

Question 28- Write a python program using RegEx to remove <U+..> like symbols Check the below sample text, there are strange symbols something of the sort <U+..> all over the place. You need to come up with a general Regex expression that will cover all such symbols.

In [35]:
```
import re

string = "Check the below sample text, there are strange symbols something of the s
string = re.sub(r'<U\+[0-9A-Fa-f]{1,4}>', '', string)
print(string)
```

```
Check the below sample text, there are strange symbols something of the sort <U+..
> all over the place. You need to come up with a general Regex expression that wil
l cover all such symbols.
```

Question 29- Write a python program to extract dates from the text stored in the text file.

In [43]:
```
import re

# open a text file
f = open("example.txt",'r')

# extract the file's content
```

```
content = f.read()

# a regular expression pattern to match dates
pattern = "\d{2}[/-]\d{2}[/-]\d{4}"

# find all the strings that match the pattern
dates = re.findall(pattern, content)

for date in dates:
    print(date)
```

```
---------------------------------------------------------------------------
FileNotFoundError                         Traceback (most recent call last)
Cell In[43], line 4
      1 import re
      3 # open a text file
----> 4 f = open("example.txt",'r')
      6 # extract the file's content
      7 content = f.read()

File ~\anaconda3\lib\site-packages\IPython\core\interactiveshell.py:282, in _modified_open(file, *args, **kwargs)
    275 if file in {0, 1, 2}:
    276     raise ValueError(
    277         f"IPython won't let you open fd={file} by default "
    278         "as it is likely to crash IPython. If you know what you are doing, "
    279         "you can use builtins' open."
    280     )
--> 282 return io_open(file, *args, **kwargs)

FileNotFoundError: [Errno 2] No such file or directory: 'example.txt'
```

Question 30- Write a Python program to replace all occurrences of a space, comma, or dot with a colon.

In [42]:
```
import re

text = 'Python Exercises, PHP exercises.'
print(re.sub("[ ,.]", ":", text))
```

```
Python:Exercises::PHP:exercises:
```

In [ ]:
```

```