In [38]:
```
%%html
<img src = "specialregex.png">
```

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 [39]:
```
import re

def check_pattern():
    result = re.findall((r'[(a-z)|(A-Z)|(0-9)]'), target_string)
    print("match object:", result)
```

In [40]:
```
target_string = "Write a Python program to check that a string contains only a certain set of characters 123-456-789"
check_pattern()
```

```
match object: ['W', 'r', 'i', 't', 'e', 'a', 'P', 'y', 't', 'h', 'o', 'n', 'p', 'r', 'o', 'g', 'r', 'a', 'm', 't', 'o',
'c', 'h', 'e', 'c', 'k', 't', 'h', 'a', 't', 'a', 's', 't', 'r', 'i', 'n', 'g', 'c', 'o', 'n', 't', 'a', 'i', 'n', 's',
'o', 'n', 'l', 'y', 'a', 'c', 'e', 'r', 't', 'a', 'i', 'n', 's', 'e', 't', 'o', 'f', 'c', 'h', 'a', 'r', 'a', 'c', 't',
'e', 'r', 's', '1', '2', '3', '4', '5', '6', '7', '8', '9']
```

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

In [41]:
```
import re

def match_string(string):
    pattern = r'ab*'
    match = re.match(pattern, string)
    if match:
        return True
    else:
        return False

print(match_string("Create a function in python abb that matches a a0 string that has an a followed by zero"))
print(match_string("abb"))
```

```
#it doesn't match abb found in the intire string because it expects all the entire string to act as the pattern;
# infact on the other with all the pattern matching the string it give me True. :)
```

```
False
True
```

In [42]:
```
import re

def match_string(string):
    pattern = r'ab*'
    match = re.search(pattern, string)
    if match:
        return True
    else:
        return False

print(match_string("Create a function in python abby that matches a a0 string that has an a followed by zero"))

# in this case it returns every word in the string that matches the pattern, good one there :)
```

```
True
```

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

In [43]:
```
import re

def match_string(string):
    pattern = r'ab+'
    match = re.search(pattern, string)
    if match:
        return True
    else:
        return False

print(match_string("Create a function in python abby that matches a a0 string that has an a followed by zero"))
```

```
True
```

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 [44]:
```
import re

def match_string(string):
    pattern = r'a(0|b)'
    match = re.search(pattern, string)
    if match:
        return True
    else:
        return False

print(match_string("I have a book"))
print(match_string("I have abnormal book"))
```

```
False
True
```

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

In [45]:
```
import re

def match_string(string):
    pattern = r'ab{3}'
    match = re.search(pattern, string)
    if match:
        return True
    else:
        return False

print(match_string("I have abnormal book"))
print(match_string("I have abbbnormal book"))
```

```
False
True
```

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 [46]:
```
import re

def split_string(string):
    pattern = r'(?=[A-Z])'
    result = re.split(pattern, string)
    return result


split_string = split_string("ImportanceOfRegularExpressionsInPython")
split_string.remove("")
split_string
```

Out[46]:
```
['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 [47]:
```
import re

def match_string(string):
    pattern = r'ab{2,3}'
    match = re.search(pattern, string)
    if match:
        return True
    else:
        return False

print(match_string("I have abnormal book"))
print(match_string("I have abbnormal book"))
print(match_string("I have abbbnormal book"))
```

```
False
True
True
```

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

In [48]:
```
import re
```

```
def find_sequences(string):
    pattern = r'[a-z]+_[a-z]+'
    sequences = re.findall(pattern, string)
    return sequences


print(find_sequences("I have abnormal_book"))
```

```
['abnormal_book']
```

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

In [49]:
```
import re

def find_sequences(string):
    pattern = r'a.*?b$'
    sequences = re.search(pattern, string)
    return sequences

print(find_sequences("I have abnormal book bob"))
```

```
<re.Match object; span=(3, 24), match='ave abnormal book bob'>
```

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

In [50]:
```
import re

def find_sequences(string):
    pattern = r'^\w+'
    sequences = re.search(pattern, string)
    return sequences


print(find_sequences("a I have abnormal book bob"))
```

```
<re.Match object; span=(0, 1), match='a'>
```

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

In [51]:
```
import re

def find_sequences(string):
    pattern_1 = r'[a-zA-Z]+'
    pattern_2 = r'[0-9]'
    pattern_3 = r'([a-zA-Z]|[0-9])+_([a-zA-Z]|[0-9]+)'

    sequences_1 = re.search(pattern_1, string)
    sequences_2 = re.search(pattern_2, string)
    sequences_3 = re.search(pattern_3, string)

    return sequences_1, sequences_2, sequences_3

# Test the function
string = "a I have abnormal 2 book bo_b"
result_1, result_2, result_3 = find_sequences(string)

print(result_1)
print(result_2)
print(result_3)
```

```
<re.Match object; span=(0, 1), match='a'>
<re.Match object; span=(18, 19), match='2'>
<re.Match object; span=(25, 29), match='bo_b'>
```

In [52]:
```
import re

def match_string(string):
    pat_1, pat_2, pat_3 = r'([a-zA-Z]+', r'[0-9]+', r'_[a-zA-Z0-9]+'
    sequences_1, sequences_2, sequences_3 = re.findall(pat_1, string), re.findall(pat_2, string),re.findall(pat_3, stri
    return sequences_1, sequences_2, sequences_3

# Test the function
string = "a I have abnormal 2 book bo_b"
result_1, result_2, result_3 = find_sequences(string)

print(result_1)
print(result_2)
print(result_3)
```

```
<re.Match object; span=(0, 1), match='a'>
<re.Match object; span=(18, 19), match='2'>
<re.Match object; span=(25, 29), match='bo_b'>
```

In [53]:
```python
import re

def find_sequences(string):
    pattern = r'([a-zA-Z]+|[0-9]+|_[a-zA-Z0-9]+)'
    sequences = re.findall(pattern, string)
    return sequences

# Test the function
string = "a I have abnormal 2 book bo_b"
results = find_sequences(string)
if results:
    print("Found sequences:", results)
else:
    print("No sequences found.")
```

```
Found sequences: ['a', 'I', 'have', 'abnormal', '2', 'book', 'bo', '_b']
```

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

In [54]:
```python
import re

def find_sequences(string):
    pattern = r'^3.*'
    sequences = re.findall(pattern, string)
    return sequences

string = "3 a I have abnormal 2 book bo_b"
print(find_sequences(string))
```

```
['3 a I have abnormal 2 book bo_b']
```

In [55]:
```python
import re

def find_sequences(string):
    pattern = r'^3(.*)'
    sequences = re.findall(pattern, string)
    return sequences
```

```
string = "3 a I have abnormal 2 book bo_b"
print(find_sequences(string))
```

```
[' a I have abnormal 2 book bo_b']
```

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

In [56]:
```
import re
IP_address = "0055-555 -556"
result = re.sub('[.0*]', "", IP_address)

print(result)
```

```
55-555 -556
```

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 [57]:
```
import re

pattern = r'\b([A-Z][a-z]+ \d{1,2}(?:st|nd|rd|th)? \d{4})\b'

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 '''

result = re.search(pattern, Sample_text)
print(result)
```

```
<re.Match object; span=(3, 19), match='August 15th 1947'>
```

In [58]:
```
import re

pattern = r'([A-Z][a-z]+ \d{1,2}(?:st|nd|rd|th)? \d{4})'

with open('C:/Users/USER/text_sample.txt', 'r') as file:
    sample_text = file.read()
```

```
result = re.search(pattern, sample_text)
print(result)
```

```
<re.Match object; span=(4, 20), match='August 15th 1947'>
```

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 [59]:
```
import re

pattern = r'(horse|dog|fox)+'

sample_text = 'The quick brown fox jumps over the lazy dog and horse.'

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

```
['fox', 'dog', 'horse']
```

In [60]:
```
import re

pattern = r'(horse|dog|fox)+'

with open('C:/Users/USER/text_sample.txt', 'r') as file:
    sample_text = file.read()

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

```
['fox', 'dog', 'horse']
```

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 [61]:
```
import re

pattern = r'fox'

sample_text = 'The quick brown fox jumps over the lazy dog and horse.'

matches = re.finditer(pattern, sample_text)
for match in matches:
    start_index = match.start()
    end_index = match.end()
    print(f"Found 'fox' at position {start_index} to {end_index-1}.")
```

```
Found 'fox' at position 16 to 18.
```

In [62]:
```
import re

pattern = r'(horse|dog|fox)+'

sample_text = 'The quick brown fox jumps over the lazy dog and horse.'

matches = re.finditer(pattern, sample_text)
for match in matches:
    start_index = match.start()
    end_index = match.end()
    print(f"Found '{match.group()}' at position {start_index} to {end_index-1}.")
```

```
Found 'fox' at position 16 to 18.
Found 'dog' at position 40 to 42.
Found 'horse' at position 48 to 52.
```

Question 17- Write a Python program to find the substrings within a string.

Sample text : 'Python exercises, PHP exercises, C# exercises' Pattern : 'exercises'.

In [63]:
```
import re

sample_text = 'Python exercises, PHP exercises, C# exercises'

pattern = r'exercises'

result = re.findall(pattern, sample_text)

print(result)
```

```
['exercises', 'exercises', 'exercises']
```

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

In [64]:
```
import re

sample_text = 'Python exercises, PHP exercises, C# exercises'
pattern = r'exercises'

matches = re.finditer(pattern, sample_text)

for match in matches:
    occurrence = match.group()
    start_position = match.start()
    end_position = match.end()

    print(f"Found '{occurrence}' at position {start_position} to {end_position - 1}.")
```

```
Found 'exercises' at position 7 to 15.
Found 'exercises' at position 22 to 30.
Found 'exercises' at position 36 to 44.
```

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

In [65]:
```
import re
date = '2022-07-15'
list = re.split("-", date)
list
```

Out[65]:
```
['2022', '07', '15']
```

In [66]:
```
new_date = "-".join(list[::-1])
print(new_date)
```

```
15-07-2022
```

In [67]:
```
# second method:

from datetime import datetime
```

```
date_str = '2022-07-15'
input_format = '%Y-%m-%d'
output_format = '%d-%m-%Y'

# Convert the string to a datetime object
date_obj = datetime.strptime(date_str, input_format)

# Convert the datetime object to a formatted string
converted_date = datetime.strftime(date_obj, output_format)

print(converted_date)
```

```
15-07-2022
```

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

In [68]:
```
import re

string = """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'
'The quick brown fox jumps over the lazy dog and horse."""

pattern = r'\b[aAeE]\w+'
result = re.findall(pattern, string )
result
```

Out[68]:
```
['August', 'and', 'and']
```

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

In [69]:
```
import re

string = "The price of the product is $25.99 and the quantity is 10."

pattern = r'\d+'

matches = re.finditer(pattern, string)
for match in matches:
```

```
    number = match.group()
    position = match.start()
    print(f"Number: {number}, Position: {position}")
```

```
Number: 25, Position: 29
Number: 99, Position: 32
Number: 10, Position: 55
```

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

In [70]:

```
import re

def max_number(string):
    pattern = r"\d+"
    numbers = re.findall(pattern, string)
    if numbers:
        return max(numbers, key=int)
    else:
        return None

string = "3 a I have 25 abnormal 2 book bo_b"
print(max_number(string))
```

```
25
```

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

In [71]:

```
import re

def add_spaces(string):
    pattern = r"([A-Z][a-z]+)"
    result = re.sub(pattern, r" \1", string)
    return result

string = "3 a I have 25 abnormal 2 book bo_b MegaGiga"
result = add_spaces(string)
print(result)
```

```
3 a I have 25 abnormal 2 book bo_b  Mega Giga
```

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

In [72]:
```
import re

def add_spaces(string):
    pattern = r"([A-Z][a-z]+)"
    result = re.search(pattern, string)
    return result

string = "3 a I have 25 abnormal 2 book bo_b MegaGiga"
result = add_spaces(string)
print(result)
```

```
<re.Match object; span=(35, 39), match='Mega'>
```

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

In [73]:
```
import re

def remove_duplicate(string):
    pattern = r'\b(\w+)\b(?=.*\b\1\b)'
    result = re.sub(pattern, '', string)
    return result

string = "3 a I 25 have 25 abnormal abnormal 2 book bo_b cool cool MegaGiga"
result = remove_duplicate(string)
print(result)
```

```
3 a I  have 25  abnormal 2 book bo_b  cool MegaGiga
```

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

In [74]:
```
import re

def alpha_num(string):
    pattern = r".*[a-zA-Z0-9]$"
```

```
    match = re.match(pattern, string)
    if match:
        return True
    else:
        return False

string = input("Enter a string: ")
if alpha_num(string):
    print("Accepted: ", "String ends with an alphanumeric character.")
else:
    print("String does not end with an alphanumeric character.")
```

```
Enter a string:
String does not end with an alphanumeric character.
```

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

In [75]:

```
import re

def extract_hashtags(string):
    pattern = r"#\w+"
    hashtags = re.findall(pattern, Sample_text)
    return hashtags

Sample_text = input("Enter a string: ")
hashtags = extract_hashtags(string)
if hashtags:
    print("Extracted hashtags:")
    for hashtag in hashtags:
        print(hashtag)
else:
    print("No hashtags found.")
```

```
Enter a string:
No hashtags found.
```

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. Sample Text: "@Jags123456 Bharat band on 28??<U+00A0><U+00BD>

<U+00B8><U+0082>Those who are protesting #demonetization are all different party leaders" Output: @Jags123456 Bharat band on 28??Those who are protesting #demonetization are all different party leaders

In [76]:
```
import re

def remove_symbols(text):
    pattern = r"<U\+[a-zA-Z0-9]+>"
    result = re.sub(pattern, "", text)
    return result

sample_text = """@Jags123456 Bharat band on 28?<U+00A0><U+00BD><U+00B8><U+0082> Those who are protesting
               #demonetization are all different party leaders"""

processed_text = remove_symbols(sample_text)
print(processed_text)
```

```
@Jags123456 Bharat band on 28?? Those who are protesting
               #demonetization are all different party leaders
```

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

Sample Text: Ron was born on 12-09-1992 and he was admitted to school 15-12-1999. Store this sample text in the file and then extract dates.

In [77]:
```
import re

pattern = r'\d{1,2}-\d{1,2}-\d{4}'


Text = """ Ron was born on 12-09-1992 and he was admitted to school 15-12-1999.
    Store this sample text in the file and then extract dates."""
dates = re.findall(pattern, Text)
dates
```

Out[77]:
```
['12-09-1992', '15-12-1999']
```

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

Sample Text- 'Python Exercises, PHP exercises.' Output: Python:Exercises::PHP:exercises:

In [78]:
```
import re

pattern = r"(\,|\.)"
Text = 'Python Exercises, PHP exercises.'
string = re.sub(pattern, ":", Text)
string.replace(" ", ":")
```

Out[78]:
```
'Python:Exercises::PHP:exercises:'
```

In [79]:
```
# second method
import re

pattern = r"[\s\.,]"
Text = 'Python Exercises, PHP exercises.'
string = re.sub(pattern, ":", Text)
string
```

Out[79]:
```
'Python:Exercises::PHP:exercises:'
```

In [ ]:
```

```