**Assignment – 7**

1. **What is the name of the feature responsible for generating Regex objects?**

Ans=> The re.compile() function returns Regex objects.

**2. Why do raw strings often appear in Regex objects?**

Ans=> Raw strings are used so that backslashes do not have to be escaped.

**3. What is the return value of the search() method?**

Ans=> The search() method returns Match objects.

**4. From a Match item, how do you get the actual strings that match the pattern?**

Ans=> The group() method returns strings of the matched text.

**5. In the regex which created from the ‘(\d\d\d)-(\d\d\d-\d\d\d\d)’, what does group zero cover? Group 2? Group 1?**

Ans=> Group 0 is the entire match, group 1 covers the first set of parentheses, and group 2 covers the second set of parentheses.

**6. In standard expression syntax, parentheses and intervals have distinct meanings. How can you tell a regex that you want it to fit real parentheses and periods?**

Ans=> Periods and parentheses can be escaped with a backslash: \., \(, and \).

**7. The findall() method returns a string list or a list of string tuples. What causes it to return one of the two options?**

Ans=> If the regex has no groups, a list of strings is returned. If the regex has groups, a list of tuples of strings is returned.

**8. In standard expressions, what does the | character mean?**

Ans=> It matches alternatives.

**9. In regular expressions, what does the character stand for?**

Ans=> A character set is an explicit list of the characters that may qualify for a match in a search. A character set is indicated by enclosing a set of characters in brackets ([ and ]).

**10.In regular expressions, what is the difference between the + and \* characters?**

Ans=> The star quantifier(\*) means that the preceding expression can match zero or more times it is like {0,}

While the plus quantifier(+) indicate that the preceding expression MUST match at least one time or multiple times and it is the same as {1,} .

**12. What do you mean by the \d, \w, and \s shorthand character classes signify in regular**

**expressions?**

Ans=> **\w:** This is the "word character" class that represents the regex range [A-Za-z0-9\_],

and it will match a single uppercase character, lower-case character, digit, or underscore.

**\d:** This is the "digit character" class represents the regex range [0-9],

and it will match the single-digit character.

**\s:** This is the "whitespace character" class that represents the regex range, matching a single space, carriage return, tab, line break, form feed, or vertical tab

**13. What do means by \D, \W, and \S shorthand character classes signify in regular expressions?**

Ans=> **\W:** the "non-word character" class represents the regex range [^A-Za-z0-9\_],

matching any character that is not included in the range represented by \w

**\D:** the "non-digit character" class represents the regex range [^0-9],

matching any character that is not included in the range represented by \d

**\S:** the “non-whitespace character” class represents the regex range [^ \t\r\n\f\v],

matching any character that is not included in the range represented by \s

**14. What is the difference between .\*? and .\* ?**

Ans=> (.\*?) matches any character (.) any number of times (\*), as few times as possible to make the regex match (?).

You'll get a match on any string, but you'll only capture a blank string because of the question mark.

This feature is much more useful when you have a more complicated regex. Here, the parser doesn't have to capture anything at all to get a match: the asterisk allows any number of characters in the capturing group, while the question mark makes the parser save as many as possible from the input text for later, resulting in nothing being captured.

(.\*) captures a group zero or one times.

That group consists of a run of any length (\*) of any character (.).

This also will match anything, but it will capture the first line, since the dot matches anything except a newline.

**15. What is the syntax for matching both numbers and lowercase letters with a character class?**

Ans=> The character class [a-zA-Z0-9] will match all lowercase letters, uppercase letters, and numbers.

**16. What is the procedure for making a normal expression in regax case insensitive?**

Ans=> Pass re.IGNORECASE to the flags param of search, match, or sub:

re.search('test', 'TeSt', re.IGNORECASE)

re.match('test', 'TeSt', re.IGNORECASE)

re.sub('test', 'xxxx', 'Testing', flags=re.IGNORECASE)

**17. What does the . character normally match? What does it match if re.DOTALL is passed as 2nd argument in re.compile()?**

Ans=> The . character matches any character, including newline. If re. DOTALL is passed as the second argument to re. compile(), then the dot will also match newline characters.

**18. If numReg = re.compile(r’\d+’), what will numRegex.sub(‘X’, ‘11 drummers, 10 pipers, five rings, 4 hen’) return?**

Ans=> import re

numRegex = re.compile(r'\d+')

numRegex.sub('X', '11 drummers, 10 pipers, five rings, 4 hen')

**O/P =>** 'X drummers, X pipers, five rings, X hen'

**19. What does passing re.VERBOSE as the 2nd argument to re.compile() allow to do?**

Ans=> The re. VERBOSE argument allows you to add whitespace and comments to the string passed to re.

**20. How would you write a regex that match a number with comma for every three digits? It must match the given following:**

**‘42’**

**‘1,234’**

**‘6,368,745’**

**but not the following:**

**‘12,34,567’ (which has only two digits between the commas)**

**‘1234’ (which lacks commas)**

Ans=> The regex is = ^\d{1,3}(?:,\d{3})\*$

**21. How would you write a regex that matches the full name of someone whose last name is Watanabe? You can assume that the first name that comes before it will always be one word that begins with a capital letter. The regex must match the following:**

**‘Haruto Watanabe’**

**‘Alice Watanabe’**

**‘RoboCop Watanabe’**

**but not the following:**

**‘haruto Watanabe’ (where the first name is not capitalized)**

**‘Mr. Watanabe’ (where the preceding word has a nonletter character)**

**‘Watanabe’ (which has no first name)**

**‘Haruto watanabe’ (where Watanabe is not capitalized)**

Ans=> The Regular Expression is - r"^[A-Z]\w+ Watanabe\n"

**22. How would you write a regex that matches a sentence where the first word is either Alice, Bob, or Carol; the second word is either eats, pets, or throws; the third word is apples, cats, or baseballs; and the sentence ends with a period? This regex should be case-insensitive. It must match the following:**

**‘Alice eats apples.’**

**‘Bob pets cats.’**

**‘Carol throws baseballs.’**

**‘Alice throws Apples.’**

**‘BOB EATS CATS.’**

**but not the following:**

**‘RoboCop eats apples.’**

**‘ALICE THROWS FOOTBALLS.’**

**‘Carol eats 7 cats.’**

Ans=> import re

regex=re.compile(r'(Alice|Bob|Carol)\s+(eats|pets|throws)\s+(apples|cats|baseballs)\.', re.IGNORECASE)