BT 3051 (July-Nov 2019) - Assignment 2

Name: N Sowmya Manojna
Roll Number: BE17B007
Collaborators: -
Time taken: 1.5 days

1. (a) **The matches made by regular expressions are greedy.**
So, it matches the first half of the '|' condition with the text provided. Once a match is found, it would automatically skip the second half of the '|' condition. In the absence of a match with the first half of the '|' condition, it would match with the second half of the '|' condition. In this case, it matches the '1' with the '1' inside the [129] and hence, skips the second half of the '|' condition. Hence, only '1' gets matched

   (b) **The regular expression states matching from the first instance where it finds a match.**
In this case, 'There.*Zero' begins matching from the first 'There' till the last 'Zero' in the text provided - 'There are 18 trunks Zeros'

2. (a) \b(?!F)\w+\b
The regular expression matches all words that doesn't start with a 'F'.

   (b) (\w+(?<!c)qu\w+)|(Qu\w+)
The regular expression matches all words that either start with a 'Qu' or have a 'qu' which isn't preceded by a 'c'.

   (c) (\b(I(n|t)i)\w+)|(\w+ini\b)
The regular expression matches all words that either start with an 'Ini' or 'Iti' or end with an 'ini'.

   (d) (?=\b\d{4,5}\b)\d+ (or) \b\d{4,5}\b
The regular expression matches all 4 or 5 digit numbers.

3. (a) Match: aeiou, acedious, azeyixowuv, abecdieofug
Mismatch: try, aioue, assignment, balance
The regular expression matches all words that have 'a', 'e', 'i', 'o' and 'u', in the same order, with or without other word character inbetween.

   (b) Match: four, more, lettered, words
Mismatch: or, try, 123, ?-*/+
The regular expression matches all words that have four or more word characters.

   (c) Match: 1231234$, 987259878$, 0129801250125$, 1111112$
Mismatch: abc$, 100201$, 01212012$, 100858745201$
The regular expression matches all numbers having a set of 3 digits, followed by an optional number of digits, the initial 3 digits with an additional digit, repeated one or more times at the end.

   (d) Match: you'll, four's, would've, could've
Mismatch: 4, lettered, words, ?
The regular expression matches all words having 1 or more word characters, followed by an apostrophe and 1 or more word characters.

   (e) Match: 1234123121$, 9876987989$, 6543654656$, 4561456454$
Mismatch: alphabets, symbols, 12345, ?+-*/
The regular expression matches all 10 digit numbers followed by a '$' symbol. In the number (without the '$'), the digits (5,6,7) are the digits (1,2,3) of the number, digits (8,9) are the digits (1,2) of the number and the digit 10 is the digit 1 of the number.

4. (a) \b(\+|\-|)\d+((\.\d+)|(\/\d+)|(d*))\b

   (b) The answer below checks for 30, 31 day variations and leap years in case of February.
   \b(((0[13578])|(1[02]))([\/\._])((0[1-9])|([12][0-9])|(3[01]))\5((0\d{2}[1-9])|([1-9]\d{3}))|
   (((0[469])|(11))([\/\._])((0[1-9])|([12][0-9])|(31))\17((0\d{2}[1-9])|([1-9]\d{3}))|
   (02([\/\._])(((0[1-9])|([12][0-9]))\26(\d{2}((04)|(08)|(12)|(16)|(20)|(24)|(28)|(32)|(36)|(40)|(44)
   |(48)|(52)|(56)|(60)|(64)|(68)|(72)|(76)|(80)|(84)|(88)|(92)|(96))|(((04)|(08)|(12)|(16)|(20)|(24)|
   (28)|(32)|(36)|(40)|(44)|(48)|(52)|(56)|(60)|(64)|(68)|(72)|(76)|(80)|(84)|(88)|(92)|(96))00))))|
   ((((0[1-9])|([12][0-8]))\1((0\d{2}[1-9])|([1-9]\d{3})))))\b

   (c) (?<!(\/\/)|(\.))\b[\w\d_]+\.[\w\d_]+$ (or)
   (?<!(\/\/)|(\.))\b[^\\\/?%*:\|"<>]+[^]\.[\w\d_]+$

   (d) ATG([ACTG]{3})*((TAA)|(TAG)|(TGA))

   (e) AUG([ACUG]{3})*?((UAA)|(UAG)|(UGA))

   (f) (?=.*[A-Z])(?=.*[a-z])(?=.*\d)(?=.*[~!#$%^&()\-_+<>?,])[A-Za-z\d~!#$%^&()\-_+<>?,]{7,15}