## BIOI 3500, Advanced Bioinformatics Programming Assignment 6



Assignment 6:: 50 points:: Due: 11:59 p.m. March, 5 2021

## Objectives

This assignment will give you the opportunity to practice using regular expressions.

## Exercises1

Use regular expressions and **grep** answer the following questions. Match against entries in the file **wordlist.txt**<sup>2</sup> (on Canvas in **Files/Assignments/Assignment06/wordlist.txt**), an eclectic collection of 224,714 words. The number in parenthesis at the end of most questions is the number of matches I found.

Record the regular expression that you used to answer to each question, the matches themselves (or the first 10 if there are more than 10), and a count of the number of matches in a text file. Name this file **netID\_Assign6.txt** where **netID** is your UNO NetID. Be sure to include your name in the file.

- 1. The currency name franc (France) is hidden in the word *efranchise* and rand (South African) is hidden in *grandma*. In what words does the currency name *dinar* (Iraq) appear in? (48)
- 2. What words contain the consecutive letters nacl? (28)
- 3. What words contain the consecutive letter combination tantan? (9)
- 4. What words start and end with a vowel (a, e, i, o, u and y) and only have one or more consonants in-between? (79)
- 5. What words contain the vowels (a, e, i, o, u and y) in order without repeats? Note that adventitiously is not a correct match because i occurs twice before o. (2)
- 6. What words begin and end with the same vowel (a, e, i, o, u, or y)? (3,180)
- 7. What words contain exactly three y's? (97)
- 8. What words contain the sequence hot or the sequence cold? (827) Do any words contain both?
- On a standard telephone keypad, each number (except 1 and 0) is associated with at least three letters (see http://en.wikipedia.org/wiki/Telephone\_keypad). Using these associations what words can the phone number 226-3385 spell? (1)
- 10. What is the longest word you can type on a standard QWERTY keyboard using only your left hand? (On a standard QWERTY keyboard the letters typed with the left hand are asdfgqwertzxcvb.) Hint: try different regular expressions with increasing size of the word until you can't find a match.
- 11. What is the longest word you can type on a standard QWERTY keyboard using only your right hand? (On a standard QWERTY keyboard the letters typed with the right hand are *hjklyuiopnm*.)
- 12. What words end in the two letters mt? (3)
- 13. What words end in dous? (125)
- 14. What is the longest word that can be typed using only those letters on the top row of a standard QWERTY keyboard? Middle row?
- 15. What words contain the letter q not immediately followed by the letter u? (6)
- 16. What words contain a sequence of three letters that is repeated elsewhere in the word? (2,554)
- 17. What words contain a sequence of three letters, the first and last of which are the same, that is repeated elsewhere in the word; e.g., the *ono* in *monotonous*? (130)
- 18. What words contain a sequence of two letters that is repeated at least three times (total) in the word? (310)
- 19. For which words are the first three letters the same as the reverse of those letters at the end? (52)
- 20. A palindrome is a word, phrase, number or other sequence of units that can be read the same way in either direction. What words are five-letter palindromes? What words are seven-letter palindromes? What is the longest palindrome in wordlist.txt?

## **How to Submit**

Submit your file as an attachment using Canvas.

<sup>&</sup>lt;sup>1</sup>Part of this assignment are adopted from *Regular Expression Play*, Wheaton College Department of Computer Science (http://cs.wheatoncollege.edu/regexplay).

<sup>&</sup>lt;sup>2</sup>Obtained from http://introcs.cs.princeton.edu/java/data