

BT3051: Assignment 3

Abhinav Sardana

BE14B002

Part I: Written

1. For each of the following regular expressions, (i) give an example of two strings that can be generated by the regular expression and two that use the same alphabet but cannot be generated, and (ii) give an English description of the set of strings generated, i. e. do not just paraphrase the regular expression operators in English; describe the sets of strings generated. In each case, the relevant alphabet/text is mentioned. Note that the special meaning characters have not been escaped as one would use with `grep -E`

(a) `'<(.)(<.)(<.)\3\2\1>'` (dictionary)

Words that match: (i) ABCDABC (ii) ABDCDBA

Words that don't match: (i) ABC (ii) AAABAAB

Description: matches 7 letter palindromes

(b) `'(.)\1.\1.\1.\1.\1'` (dictionary)

Words that match: (i) ABAACADA (ii) BABBADBAB

Words that don't match: (i) ABC (ii) AABAAB

Description: matches words that have one character 5 times or more

(c) `'AUG[AUCG]3+(UGA—UAA—UAG)'` (RNA)

Words that match: (i) AUGACGAGCUAG (ii) AUGAGCAGCUGA

Words that don't match: (i) AGCAGCAGC (ii) AUGAGCAUG

Description: RNA that begin with the start codon and end with a stop codon with atleast 1 codon between the two

(d) `'^([^\s]+)\1$'` (text)

Words that match: (i) radar is radar (ii) hat ,a red hat

Words that don't match: (i) rap ar (ii) radar

Description: A sentence that ends with the same set of letters before the first space

(e) `'(0—\+91)-*[987][0-9]4-*[0-9]5'` (text)

Words that match: (i) 09003682142 (ii) +91-90036-82142

Words that don't match: (i) +91-9003-682142 (ii) 9003682142

Description: A mobile number in the format.

(f) `'MC?M(XC—XL—L?X0,3)(IX—IV—V?I0,3)'` (text) Words that match:

(i) MM (ii) MCMLXXXVIII

Words that don't match: (i) XLI (ii) MCXLI

Description: Roman numerals in a certain range.

(g) `'<([A-Z]+)>[^\<]+\</\1>'` (HTML)

Words that match: (i) `<TITLE>BT3051</TITLE>` (ii) `<BODY>empty website</BODY>`

Words that don't match: (i) `<TITLE>BT3051</title>` (ii) `<BODY>empty<website</BODY>`

Description: HTML start and end tag in capital letters with everything in between but there should be no <.

(h) `'#[A-Za-f0-9]6'` (HTML)

Words that match: (i) `#000000` (ii) `#ffff`

Words that don't match: (i) `#00000` (ii) `#ffff`

Description: HTML colours in 6 character hexadecimal preceded by a hash.