Perl Regular Expressions

7th Feb 2015

1. Write a Perl Regular expression to test whether a string looks like a valid IP address.
2. Construct a regular expression that matches:
   1. at least one a followed by any number of b 's
   2. any number of backslashes followed by any number of asterisks (any number might be zero)
   3. three consecutive copies of whatever is contained in $whatever
   4. any five characters, including newline
   5. the same word written two or more times in a row (with possibly varying intervening whitespace), where "word" is defined as a nonempty sequence of nonwhitespace characters
3. Write a Perl program using regular expressions to create a file containing intron number, gene names, and chromosome number in three tab delimited columns extracted from the command line from the fasta sequences within /home/afedorov/EID/hs37.intEID in inton.

For example:

> INTRON\_**2** 1\_NT\_077402 protein\_id:XP\_002342051.1; Homo sapiens chromosome **1** genomic contig, GRCh37 reference primary /gene="**LOC100287102**"; intron(phase:20,size:367,681,intr\_sum:10

48); exon(size:354,127,1007,ex\_sum:1488); {splice:gtag,gtag}; CDS\_start=317, CDS\_end=1766, CDS\_len=402

2 1 LOC100287102

1. From the above file created in Question 3, write a program which creates a file with all the genes sorted in alphanumeric order of gene names followed by chromosome number and intron number.
2. Write a script that asks users for their name, email address and phone number. Test each input for accuracy, for example, there should be no letters in a phone number (let's ignore that that's not always true in this country). A phone number should have a certain length. An address should have a certain format, etc. Ask the user to repeat the input in case your script identfies it as incorrect.