

Thank you



I hope that you are
completely satisfied

Thank you Steve for using my services. I truly hope that I have created a solution to your problem.

Your Specification

Hi, i need a script to extract information from text files. I don't care how it's done - so long as it's easily available / usable in linux - python, perl, sed/awk ... up to you to decide which allows you to most easily solve the problem.

I've attached three example files. These are aligning sequences of letters and using symbols under them to represent identical (*) very similar (:) weakly similar (.) and dissimilar (no symbol / blank). Each letter in the sequence has a number. What i want to do is to list the residue numbers in groups that are identical, very similar, weakly similar, and dissimilar - using the numbering for the top sequence. I'd like the numbers listed with a + symbol between them and no line breaks or spaces etc...

In the test0 file, the text with VAPA_HUMAN on the left of the sequence is sequence name information - the 1st 65 residues of the 249 would be interpreted and grouped in a new text file as follows:

dissimilar: 1+2+3+4+5+6+7+8+9+10+11+12+13+14+15+17+18+19+21+22+24+32+33+35+36+37+40+42+58

weakly similar:

27+38+46+48+

very similar:

25+29+31+39+45+47+57+65

identical:

16+20+23+26+28+30+34+41+43+44+49+50+51+52+53+54+55+56+59+60+61+62+63+64

Results

The program parses all the data 100%.

The command line parameters are as follows:

```
python parser.py -input [inputfile] -output [outputfile]
```

I have run a test and I have included the output files for each of your example input files.

All code is commented

Final word

If you need any assistance in the future or follow up work, you are more than welcome to contact me.

Also, please remember to add your review about my work – this will help me to help others because good reviews help employers to find quality workers.

Take care.