Name: Neha Kanojiya

Class- MSc CS – I

Roll No.- 536

Subject – Bioinformatics

Topic – Regular Expression

Practical No: 8

Aim: Generate a regular expression enter three protein sequence of three different organism. Write Python/Java code to find regular expression for this sequences.

Code:

```
def gen_reg_exp(seq_list, no_of_col):
   final_list=[]
   for colnum in range(no_of_col):
      collist=[]
      for colseq in seq_list:
         collist.append(colseq[colnum])
      if len(set(collist))==len(collist):
         #print(final_list)
         final_list.append('x')
      else:
         if len(set(collist))==1:
            final_list.append(collist[0])
         else:
            final_list.append(''.join(set(collist)))
   display_output(final_list)
def display_output(final_list):
   print(*final_list, sep='-')


no_of_seq=int(input("Enter the number of sequence: "))
print("Enter all the sequences")
seq_list=[]
for _ in range(no_of_seq):
   seq_list.append(list(map(str, input("").split())))
gen_reg_exp(seq_list, len(seq_list[0]))
```

Output:


IDLE Shell 3.10.5
File Edit Shell Debug Options Window Help
Python 3.10.5 (tags/v3.10.5:f377153, Jun 6 2022, 16:14:13) [MSC v.1929 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
========== RESTART: C:/Users/DELL/Desktop/neha folders/P8.py ==========
Enter the number of sequence: 4
Enter all the sequences
A D L G A V F A L C D R Y F Q
S D V G P R S C F C E R F Y Q
A D L G R T Q L R C D R Y Y Q
A D I G Q P H S L C E R Y F Q
AS-D-IVL-G-x-x-x-x-RLF-C-ED-R-YF-YF-Q
>>>
Ln: 12 Col: 0