

BIOINFORMATICS

LAB ASSIGNMENT 2

Longest Common Subsequence

Submitted by

Nitesh Yadav
BT18CSE045

PYTHON

```
subseq.py > ...
1  from difflib import SequenceMatcher
2  str_a = open("/home/nik/cp/str1.txt")
3  str_b = open("/home/nik/cp/str2.txt")
4  str_a = str_a.read()
5  str_b = str_b.read()
6
7  def lcs(s1, s2):
8      matrix = [["" for x in range(len(s2))] for x in range(len(s1))]
9      for i in range(len(s1)):
10         for j in range(len(s2)):
11             if s1[i] == s2[j]:
12                 if i == 0 or j == 0:
13                     matrix[i][j] = s1[i]
14                 else:
15                     matrix[i][j] = matrix[i-1][j-1] + s1[i]
16             else:
17                 matrix[i][j] = max(matrix[i-1][j], matrix[i][j-1], key=len)
18             cs = matrix[-1][-1]
19         return len(cs), cs
20     print(lcs(str_a, str_b))
```

PROBLEMS OUTPUT **TERMINAL** GITLENS DEBUG CONSOLE

```
nik ~ | cp python3 subseq.py
(573, 'ACCGTTTGGTCAACACTCAACGCGCGACCGTAAAGATGGTTTTACAACCCCTCCCTGTCCGGTTTATTTCTCTAGTCAGGACAAATAAG
GAATCAGGTTAACCTATGCACCTACTTGCCGATCCAGTGAACGACTATTGTACCATCGCGTAAAGGCTAGTCGATATCCTCCAGCTAGCTTCT
```

Advantages	Disadvantages
<ol style="list-style-type: none">1. Problem-solving skills are enhanced2. Learning dynamic programming	<ol style="list-style-type: none">1. Fully Coding the approach is tough2. Learning the approach and dynamic programming is time-consuming

R

The screenshot shows the RStudio environment. The script editor contains the following code:

```

1 library(qualV,stringi,readr)
2
3 str__a = read_file("D:/SEM8/BIOINFOMATICS/Lab/Lab2/str1.txt")
4 str__b = read_file("D:/SEM8/BIOINFOMATICS/Lab/Lab2/str2.txt")
5
6 print(nchar(str__a))
7 print(nchar(str__b))
8 str_a = strsplit(str__a, split=NULL)[[1]]
9 str_b = strsplit(str__b, split=NULL)[[1]]
10 result = (LCS(str_a,str_b))
11 str = result[4]
12 typeof(str)
13 str <- stri_join_list(str,sep="")
14 print(str)
15 print(nchar(str))
16 cat(str)
17
18 |
  
```

The console output shows the execution of the last few lines of the script:

```

> typeof(str)
[1] "list"
> str <- stri_join_list(str,sep="")
> print(str)
[1] "ACCGCAGCGCCCAATTTACAACGCCGCACCGTAAAGTGGTTTTACAACCCCTCCTTGCCGGTTTATTTCTCTAGTCAGGTACAAAACAATGGTGGGAATACTTTCT
  
```

Advantages	Disadvantages
<ol style="list-style-type: none"> 1. Vast Variety of Libraries 2. Learned about File I/O operations in R Learned about the modules & functions such qualV, LCS 	<ol style="list-style-type: none"> 1. Finding the exact packages takes is time-consuming in the R documentation 2. Limited resources available online