

Lab Session 10

14.10.2024

BT3051 - DSA Biology Lab

Problem 1: Write regular expressions for the following cases: (Remember the problem from Assignment 1 where we wanted to check if the given PAN ID/Roll No is valid?)

- A valid PAN ID. Examples: AAAAAA1111D, BCBCA2019E
- A Valid Student roll no. Examples: BT20D007, CY21M101
- For the student roll numbers, match only the ones that are from the past 10 years (2015 - 2024) and also if the program is valid (5th char should be from (B, M, S, D, Z))
- A password with minimum 8 characters, at least one upper case and lower case letters, and at least one digit. No special characters are allowed.

Write a function that checks the validity of each item.

Problem 2:

Read the given FASTA file (`BT3051_lab10_Q2.fasta`), and write Regular expression to retrieve the Protein ID and the organism name for each protein. Also, extract only the sequences from the file.

Problem 3:

From the sequences extracted, find the following motifs:

- N-linked glycosylation sites: Asparagine - Any amino acid except Proline - Serine or Threonine. Examples: NNS, NAT, NLS
- Proline-rich motif: 3 or more Prolines. Examples: PPP, PPPPP