

NLP APLICADO A DADOS GENÉTICOS

Descrição do projeto:

A procura de vias de sinalização ou da função a que se encontram ligadas as proteínas é um passo crucial na interpretação de dados genéticos. Na análise de dados genómicos procura-se determinar quais os genes envolvidos numa dada patologia, criando desta forma um perfil genético. Apesar deste passo ser de grande importância o seu valor é bastante reduzido no caso de não haver interpretação biológica sobre a codificação associada aos genes. Tipicamente os investigadores depois de selecionados os genes procuram manualmente as vias de sinalização que eles representam e as proteínas que codificam e a sua função. Este passo é muito trabalhoso e moroso aumentando a dificuldade com o número de genes selecionados. A opção é então automatizar este passo, o que pode ser feito com recurso a processamento de linguagem natural (NLP).

Objetivos:

O objetivo deste estágio é implementar um pipeline de NLP para procura de vias de sinalização e função proteómica.



Francisco Caramelo
Bárbara Oliveiros Ilda Ribeiro **FMUC, iCBR**

NLP APLICADO A DADOS GENÉTICOS

Área de Investigação/ Conceitos:

- NLP;
- genómica;
- proteómica

Requisitos do estudante:

- Saber programar;
- Ter noções de estatística;

Data de inicio e Carga Horária:

- Data de início: 01/03/2022
- Carga horária: 8 semanas + 4h por semana

Equipamento Necessário:

- Computador com Python, Matlab ou R

Francisco Caramelo

Bárbara Oliveiros Ilda Ribeiro

FMUC, iCBR