

## Trabalho 1 – parte 1: CRF para NER

O reconhecimento de entidade nomeada é a tarefa de identificar entidades nomeadas de certos tipos em textos. Por exemplo:

Ministro	O
Luis	B-PESSOA
Felipe	I-PESSOA
Salomão	I-PESSOA
,	O
Quarta	B-ORGANIZACAO
Turma	I-ORGANIZACAO
,	O
julgado	O
em	O
14/11/2017	B-TEMPO

No exemplo, as etiquetas NER são fornecidas em um esquema de tags BIO. Três entidades nomeadas estão presentes: **Luis Felipe Salomão, Quarta Turma e 14/11/2017.**

As tags O indicam tokens que não fazem parte de uma entidade nomeada; as tags B indicam o início de uma entidade nomeada e as tags I indicam a continuação da entidade nomeada anterior.

As tags B e I são hifenizadas e contém um tipo após o hífen, que neste exemplo são PESSOA, ORGANIZAÇÃO e TEMPO.

### Sua tarefa

1. Alterar o sistema (CRF para NER) disponível em:  
<https://colab.research.google.com/drive/116NEMO5ZFENDk957iTVEhfxlc2EgNXNr>, criando novas features.
2. Treinar e Testar o modelo criado com o conjunto de dados fornecido (Dataset – Leis e Tweets) em português.

### Referências

<https://www.depends-on-the-definition.com/named-entity-recognition-conditional-random-fields-python/>

[https://eli5.readthedocs.io/en/latest/tutorials/sklearn\\_crfsuite.html](https://eli5.readthedocs.io/en/latest/tutorials/sklearn_crfsuite.html)

<https://towardsdatascience.com/named-entity-recognition-and-classification-with-scikit-learn-f05372f07ba2>