



**Data Science
Academy**

www.datascienceacademy.com.br

Processamento de Linguagem Natural

Classes NER



As ferramentas NER geralmente segregam as entidades em algumas classes predefinidas. Diferentes ferramentas NER possuem diferentes tipos de classes. A ferramenta Stanford NER (<https://nlp.stanford.edu/software/CRF-NER.html>) por exemplo, possui três versões diferentes baseadas nas classes NER: a primeira versão é a ferramenta NER de três classes que pode identificar as entidades - seja local, pessoa ou organização. A segunda versão é a ferramenta NER de quatro classes que pode identificar o Local, Pessoa, Organização e Misc. Misc é referido como um tipo de entidade miscelânea. Se uma entidade não pertence a Local, Pessoa ou Organização e ainda é uma entidade, então você pode marcá-la como Misc. A terceira versão é uma ferramenta de sete classes que pode identificar Pessoa, Local, Organização, Dinheiro, Porcentagem, Data e Hora. Outro exemplo de ferramenta NER é o analisador spaCy (<https://spacy.io/>) para Python, que também possui um pacote NER disponível com as seguintes classes:

- PERSON identifica o nome de uma pessoa
- NORP que significa grupos de nacionalidade, religiosos ou políticos
- FACILITY, incluindo edifícios, aeroportos, rodovias e assim por diante
- ORG para organização, instituição e assim por diante
- GPE para cidades, países e assim por diante
- LOC para locais não pertencentes à GPE, como cadeias de montanhas e oceanos
- PRODUTO que inclui objetos, veículos, alimentos e assim por diante, mas não serviços
- EVENT classe para eventos esportivos, guerras, chamados furacões, e assim por diante
- WORK_OF_ART classe para títulos de livros, músicas, e assim por diante
- LANGUAGE que marca qualquer idioma nomeado

Além disso, o pacote NER da spaCy possui classes, como data, hora, porcentagem, dinheiro, quantidade, ordinal e cardinal.

O spaCy suporta o Português. Vejamos um exemplo na próxima aula.