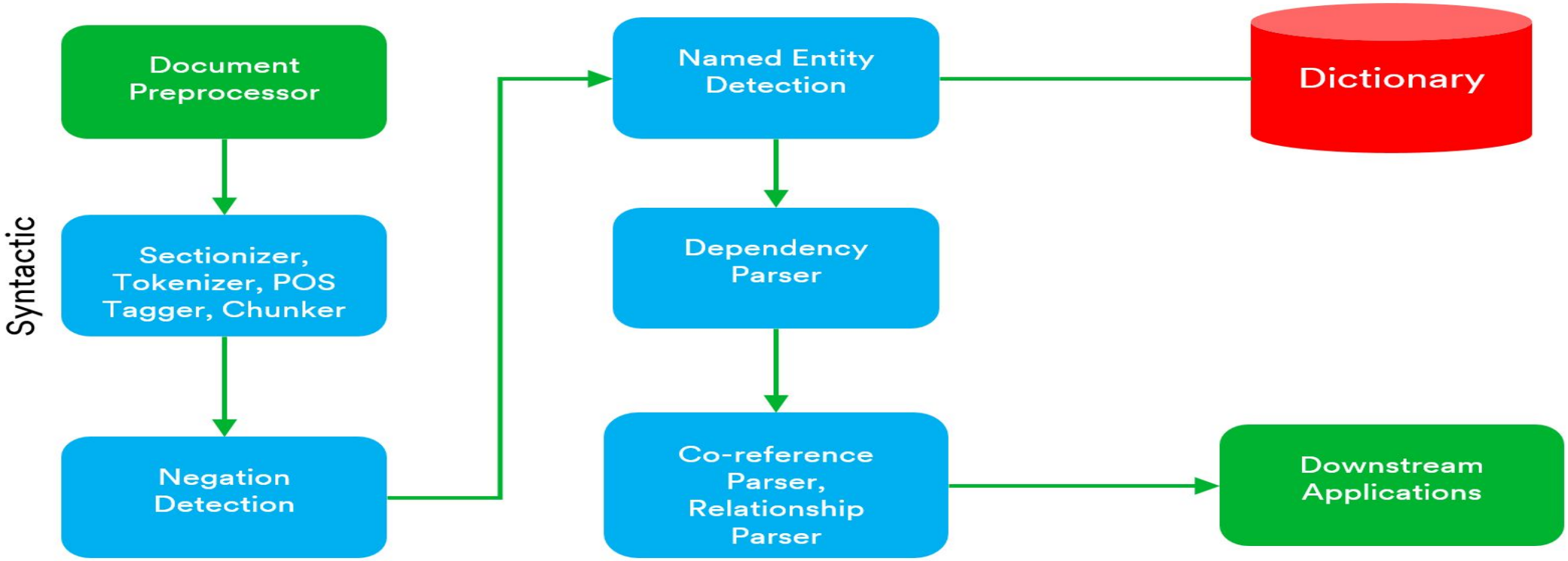
Professor(a):
José Wellington Franco da Silva
wellington@crateus.ufc.br

Análise de Dados Textuais

ANO:2019

Pipeline de processamento de texto

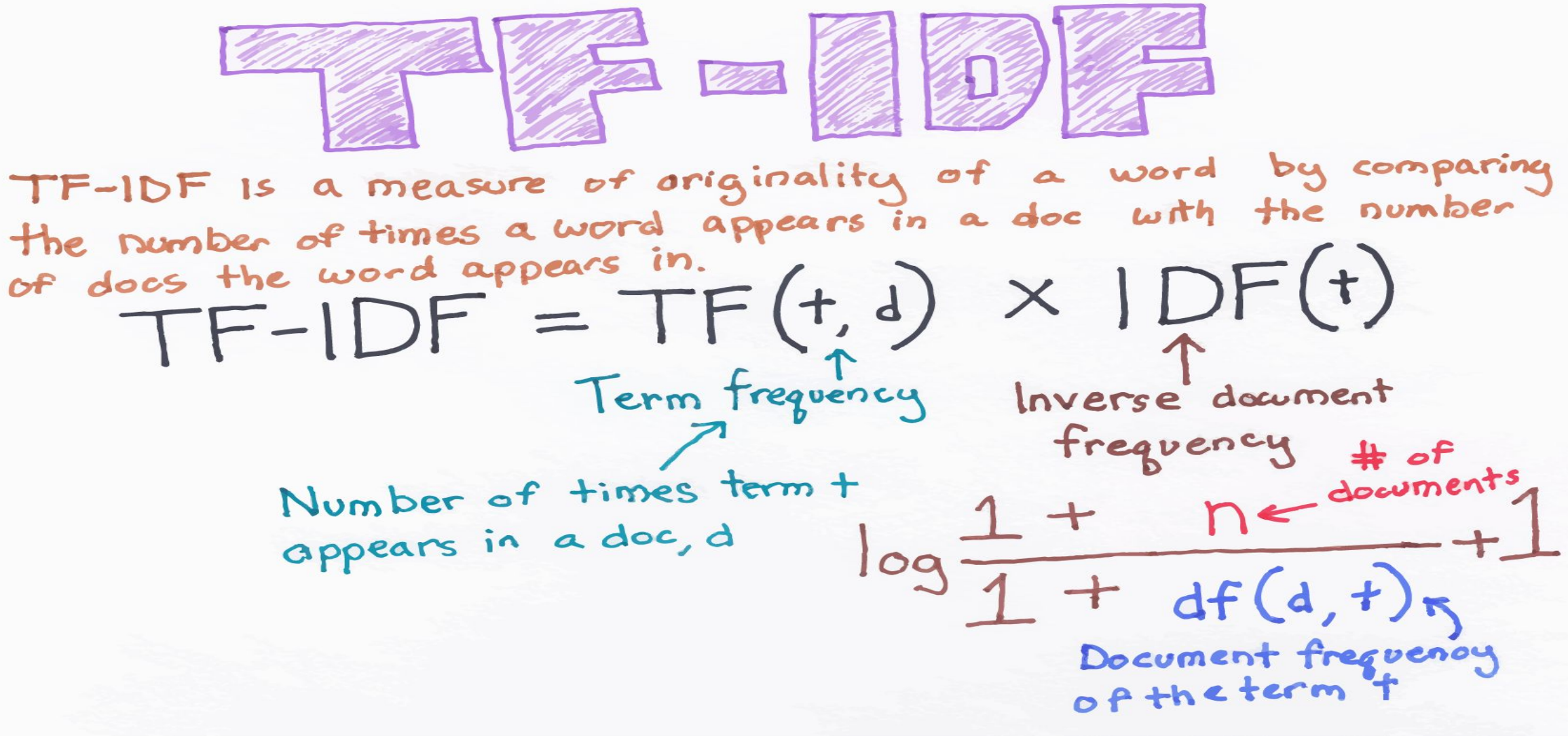
Typical Components of an NLP Application



Term Frequency, Inverse Document Frequency

- **tf–idf (abreviação do inglês *term frequency–inverse document frequency*, que significa frequência do termo–inverso da frequência nos documentos), é uma medida estatística que tem o intuito de indicar a importância de uma palavra de um documento em relação a uma coleção de documentos ou em um corpus linguístico;**
- **Tf–idf de uma palavra aumenta proporcionalmente à medida que aumenta o número de ocorrências dela em um documento, no entanto, esse valor é equilibrado pela frequência da palavra no corpus**

Term Frequency, Inverse Document Frequency



Dúvidas são bem vindas!!

- wellington@crateus.ufc.br