

Professor(a):

José Wellington Franco da Silva

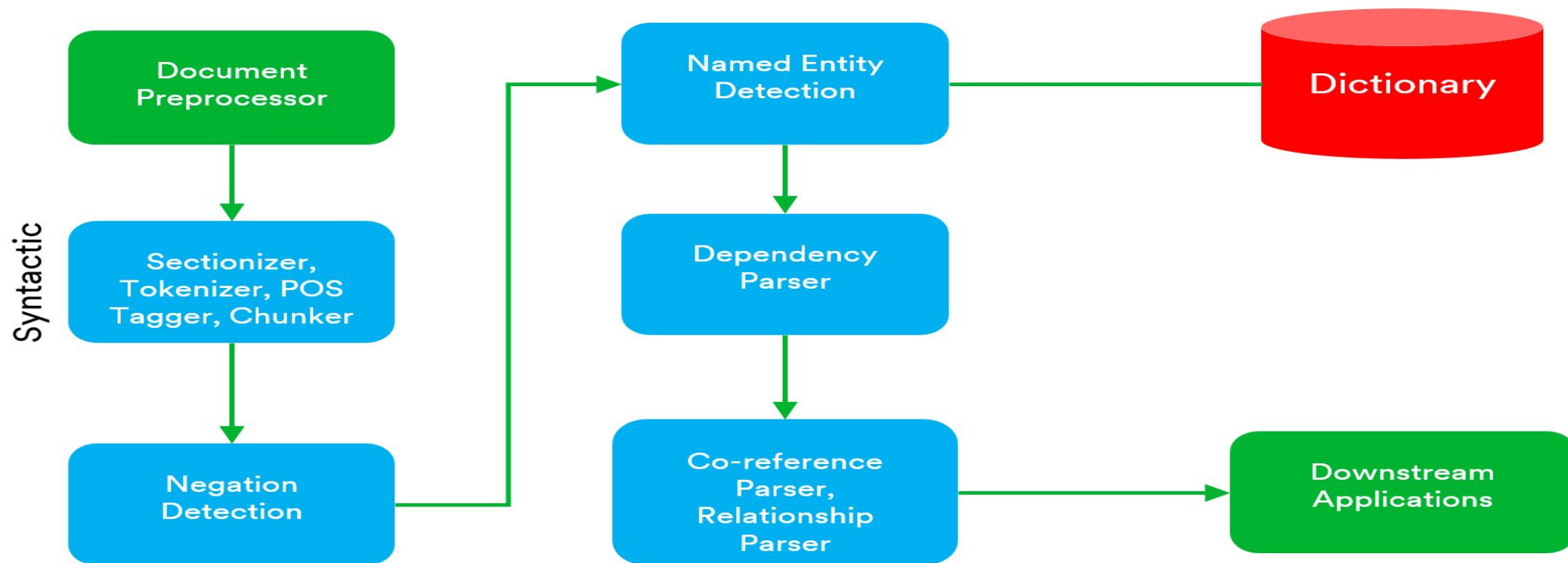
wellington@crateus.ufc.br



Análise de Dados Textuais

ANO:2019

Typical Components of an NLP Application



Nem todas as palavras são iguais...

- **Substantivos normalmente são as palavras mais significativas de um texto.**
- **Artigos têm pouca influência no significado de um texto**

- Na computação, uma palavra vazia (ou **stop word**, em inglês) é uma palavra que é removida antes ou após o processamento de um texto em linguagem natural;
- Palavras muito frequentes que acrescentam pouca semântica;
- São consideradas palavras não funcionais ;
- Normalmente são ignoradas pelos motores de busca;

Exemplos de stopwords

- <https://gist.github.com/alopes/5358189>

Dúvidas são bem vindas!!

- wellington@crateus.ufc.br



