

***INTELIGÊNCIA ARTIFICIAL***

**UA02 / LABORATÓRIO # 6**

**PLN (Processamento de Linguagem Natural)**

Em Processamento de Linguagem Natural (PLN), os seguintes termos são fundamentais:

**1.) Stopwords**

**Definição:** Palavras muito frequentes em um idioma que geralmente não carregam significado relevante para análise de texto (ex.: "o", "e", "de", "em").

**Objetivo:** Removê-las para reduzir ruído e focar em termos significativos.

**2.) Normalização**

**Definição:** Processo de padronizar palavras para uma forma canônica (base).

Técnicas incluem:

**Stemming:** Redução de palavras a seu radical (ex.: "correndo" → "corr").

**Lematização:** Transformação para o lema (forma de dicionário, ex.: "correndo" → "correr").

**Minúsculas:** Converter todo o texto para letras minúsculas.

**3.) Tokenização**

**Definição:** Dividir um texto em unidades menores (tokens), como palavras, frases ou símbolos.

Exemplo: "Olá, mundo!" → ["Olá", ",", "mundo", "!"].

**4.) Chunks (Pedaços ou Fragmentos)**

**Definição:** Segmentos de texto agrupados por critérios sintáticos/semânticos, como frases nominais (ex.: "o gato preto").

**Objetivo:** Identificar unidades linguísticas relevantes para tarefas como extração de informações.

**Resumo das Aplicações:**

- Stopwords e normalização preprocessam o texto.

- Tokenização quebra o texto em partes.

- Chunks agregam tokens para análise estrutural.

Esses processos são essenciais para modelos de PLN como classificadores, tradução automática e chatbots.