

**Curso:** DSM

**Disciplina:** Processamento de Linguagem Natural

**Turno:**

**Tipo de Avaliação:**

**Matrícula:** \_\_\_\_\_

**Aluno:** \_\_\_\_\_

**Professor:** Vagner Macedo

**NOTA**

--

## Prova de Programação de Linguagem Natural

### Instruções:

- Responda a todas as questões.
- Questões de múltipla escolha devem ser respondidas marcando a alternativa correta.
- Questões dissertativas devem ser respondidas com clareza e detalhes.

### Questões de Múltipla Escolha:

1. Qual das seguintes bibliotecas é frequentemente usada para processamento de linguagem natural (NLP) em Python?
  - a) Matplotlib
  - b) TensorFlow
  - c) NumPy
  - d) NLTK
2. Qual é a finalidade do pré-processamento de texto em NLP?
  - a) Reduzir o tamanho do texto
  - b) Melhorar a legibilidade do texto
  - c) Preparar o texto para análise NLP
  - d) Traduzir o texto para outro idioma
3. Em NLP, o que é um "stemmer"?
  - a) Um algoritmo para remover caracteres especiais de um texto.
  - b) Um algoritmo para converter palavras em sua forma básica ou radical.
  - c) Um algoritmo para traduzir texto para outro idioma.
  - d) Um algoritmo para gerar resumos automáticos.
4. Qual das seguintes tarefas não é uma aplicação típica de NLP?
  - a) Classificação de sentimentos
  - b) Tradução automática de idiomas
  - c) Reconhecimento de padrões em imagens
  - d) Extração de informações de textos

### Questões Dissertativas:

6. Explique o processo de tokenização em NLP. Dê um exemplo de uma frase tokenizada.
7. ~~Como funcionam os modelos de linguagem pré-treinados em NLP, como o GPT-3? Quais são algumas aplicações práticas para esses modelos?~~
8. Descreva a diferença entre o stemming e o lematização em pré-processamento de texto em NLP. Dê exemplos de palavras após cada um desses processos.
9. Explique o que é o processo de "embedding" de palavras (word embedding) em NLP e como ele é usado em tarefas de análise de sentimentos.

10. Suponha que você está construindo um programa para resumir textos de atendimento ao cliente. Descreva os passos principais que você seguiria para projetar e treinar um modelo de NLP que possa responder a perguntas dos clientes de forma eficaz.

### Questão Prática

11. Limpar e pré-processar um conjunto de textos antes da análise de sentimentos. O texto deverá ser um tupla com várias sentenças.
- Remova pontuações e emojis dos textos.
  - Converta todo o texto para minúsculas.
  - Tokenize cada frase em palavras.
  - Remova stopwords da língua portuguesa usando `nltk.corpus.stopwords`.
12. Escolher um conjunto de frases em arquivo csv:
- Separe as frases em duas categorias: **positivas** e **negativas**.
  - Extraia a frequência de palavras em cada categoria usando `nltk.FreqDist`.
  - Liste as 5 palavras mais comuns em cada categoria.
13. Um estudante de PLN escreveu o seguinte código em Python usando a biblioteca SpaCy. O objetivo era tokenizar o texto e imprimir apenas os tokens que NÃO são stop words. No entanto, o código contém um erro que impede que ele funcione corretamente. Identifique a linha de código que precisa ser corrigida.
- `if not token.is_stop:`
  - `if not token.text:`
  - `if token in nlp.is_stop:`
  - `import spacy`
14. Considere o texto: 'Os cientistas pesquisam muito para curar doenças.' Qual das opções a seguir representa a tokenização correta e a remoção das stop words usando NLTK?
- `['os', 'cientistas', 'pesquisam', 'curar', 'doenças']`
  - `['cientistas', 'pesquisam', 'curar', 'doenças']`
  - `['cientistas', 'pesquisam', 'muito', 'curar', 'doenças']`
  - `['Os', 'cientistas', 'pesquisam', 'muito', 'para', 'curar', 'doenças', '.']`
15. Qual é a principal diferença entre Lematização (com SpaCy) e Stemização (com NLTK) para a palavra 'andando'?
- Lematização retorna 'andar', enquanto Stemização retorna 'andand'.
  - Lematização retorna 'andou', enquanto Stemização retorna 'andando'.
  - Lematização retorna 'andar', enquanto Stemização retorna 'and'.

D. Lematização retorna 'andando', enquanto Stemização retorna 'and'.

16. Em um pipeline de PLN, por que a remoção de stop words é geralmente realizada após a tokenização?

- A. Porque a lista de stop words é baseada em tokens, não no texto completo.
- B. Porque a tokenização e a remoção de stop words não têm uma ordem específica e podem ser realizadas de forma intercambiável.
- C. Porque a remoção de stop words aumenta a precisão da tokenização.
- D. Porque a remoção de stop words é mais precisa quando o texto está na sua forma bruta, sem ser dividido em tokens.

17. Qual dos seguintes cenários seria mais apropriado para usar a stemização em vez da lematização?

- A. Em um sistema de recuperação de informações onde a velocidade é mais importante do que a precisão gramatical.
- B. Para a análise de sentimentos em mídias sociais, onde o contexto da palavra é fundamental.
- C. Em um modelo de sumarização de texto que precisa manter a fluidez e a legibilidade do conteúdo original.
- D. Ao construir um chatbot que precisa de respostas gramaticalmente corretas

18. Um desenvolvedor está usando a biblioteca TextBlob-pt para analisar sentimentos. Ele aplica a função de polaridade em duas frases: 'Adorei o filme!' e 'O filme é bom, mas o final foi ruim.'. Qual o resultado mais provável da polaridade de cada frase, respectivamente?

- A. Ambas as frases terão polaridade negativa.
- B. Ambas as frases terão polaridade neutra.
- C. A primeira frase terá polaridade positiva, e a segunda terá polaridade neutra.
- D. A primeira frase terá polaridade positiva, e a segunda terá polaridade negativa.