



Plano de Disciplina

Identificação				
CURSO Mestrado Profissional em Tecnologia da Informação			CAMPUS João Pessoa	
DISCIPLINA Tópicos Avançados em Ciência de Dados: PLN e Aplicações baseadas em Modelos de Linguagem			CÓDIGO DA DISCIPLINA POS.0330	
PRÉ-REQUISITO Não há pré-requisito				
UNIDADE CURRICULAR (OBRIGATÓRIA, OPTATIVA, ELETIVA) Optativa			SEMESTRE 2025.2	
DOCENTES RESPONSÁVEIS Damires Yluska de Souza Fernandes Alex Sandro da Cunha Rego				
Carga Horária				
TEÓRICA 32	PRÁTICA 20	EaD ¹ 8	CARGA HORÁRIA SEMANAL: 4	CARGA HORÁRIA TOTAL 60

Ementa

Fundamentos básicos sobre Processamento de Linguagem Natural (PLN); Paradigmas de PLN; Pré-processamento de texto; Arquiteturas modernas em PLN; Modelos de Linguagem Pré-Treinados (LLM); Aplicações de PLN: do clássico ao moderno; Métricas de avaliação em PLN.

Objetivos

Prover uma visão geral sobre o Processamento de Linguagem Natural e estratégias de pré-processamento de texto; conhecer e explorar tarefas clássicas de PLN; estudar arquiteturas modernas baseadas em Transformers e LLMs; identificar desafios para pesquisa e inovação em PLN por meio da discussão de problemas em aberto e possibilidades de soluções em problemas práticos advindos da indústria e da academia em cenários atuais. Desenvolver aplicações práticas integrando PLN a problemas reais.

Conteúdo Programático

Tópico	Cargahorária
1. Fundamentos de Processamento de Linguagem Natural	4
<ul style="list-style-type: none">Definição e histórico de PLN;Morfologia, sintaxe e semântica;Abordagens no PLN: simbólica, estatística, neuralDesafios computacionais	
2. Processamento de Texto	8
<ul style="list-style-type: none">TokenizaçãoNormalizaçãoLematização	

¹ Para a oferta de disciplinas na modalidade à distância, integral ou parcial, desde que não ultrapasse 20% (vinte por cento) da carga horária total do curso, observar o cumprimento da Portaria MEC nº 1.134, de 10 de outubro de 2016.



<ul style="list-style-type: none">• Stemming• Stopwords• Algoritmo Byte-Pair Encoding (BPE)• Dicionários léxicos• Representações: Bag-of-Words (BoW) e TF-IDF• Word Embeddings: Word2Vec, GloVe	
3. Datasets e Corpora	4
<ul style="list-style-type: none">• Definição• Construção de datasets• Anotação manual e automática: desafios de qualidade, custo e viés• Características de um bom dataset	
4. Tarefas de PLN	4
<ul style="list-style-type: none">• Reconhecimento de Entidades Nomeadas (Named Entity recognition)• Análise de Sentimento• Classificação de Texto• Geração de Linguagem Natural	
5. Modelos Neurais e Transformers	16
<ul style="list-style-type: none">• Introdução às Redes Neurais Profundas• Sequence to sequence (Seq2Seq)• Arquitetura encoder-decoder• Mecanismo de atenção• Modelos neurais pré-treinados: BERT, GPT, LLaMa• Modelos de Linguagem em Larga Escala (LLMs)• Fine-tuning e transfer learning• Repositório HuggingFace	
6. Abordagem de suporte ao uso de LLMs	12
<ul style="list-style-type: none">• Engenharia de Prompt• Técnica RAG (Retrieval-Augmented Generation)• Data Augmentation	
7. Avaliação Quantitativa e Qualitativa no uso de LLMs	12
<p>Quantitativa</p> <ul style="list-style-type: none">• Geração de Texto: perplexity, Bleu, Rouge, BERTscore• Classificação: Acurácia, F1-score, precisão, revocação• Ragas Metrics <p>Qualitativa</p> <ul style="list-style-type: none">• Human-in-the-loop• LLM as a judge (qualitativa automatizada)	
Total	60

Metodologia de Ensino

Aulas teóricas e práticas.



Análises e discussões sobre cases, pesquisas e soluções no escopo da disciplina.

Planejamento e execução de projeto prático envolvendo tópicos do conteúdo programático.

Reuniões de acompanhamento dos projetos para cada time de desenvolvimento e entrega de milestones em datas pré-estabelecidas.

Recursos Didáticos

Ferramentas de software open source, periódicos, artigos, Google Classroom, laboratório, acesso à Internet.

CrITÉrios de Avaliação

A avaliação dar-se-á com base em três marcos: (i) Engajamento do aluno ao longo da disciplina (pesquisas, exercícios, apresentações), (ii) Atividades práticas pontuadas, e (iii) Projeto e implementação de aplicação de PLN, com demonstração e avaliação do produto construído e apresentação oral do trabalho. Todas as atividades terão submarcos e deadlines definidos a serem cumpridos.

Bibliografia

Básica

Caseli, H.M.; Nunes, M.G.V. (org.) **Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português**. 2 ed. BPLN, 2024. Disponível em: <https://brasileiraspln.com/livro-pln/2a-edicao>.

JURAFSKY, Daniel; MARTIN, James H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models. 3. ed. Online manuscript released August 24, 2025. [S.l.]: [s.n.], 2025. Disponível em: <https://web.stanford.edu/~jufafsky/slp3/>. Acesso em: 2 set. 2025.

McShane e Nirenburg 2021 MCSHANE, M.; NIRENBURG, S. Linguistics for the Age of AI. The MIT Press, 2021. ISBN 9780262363136. Disponível em: <<https://doi.org/10.7551/mitpress/13618.001.0001>>

Complementar

VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N.; KAISER, Łukasz; POLOSUKHIN, Illia. Attention is all you need. arXiv preprint arXiv:1706.03762, 2017. Disponível em: <https://arxiv.org/abs/1706.03762>. Acesso em: 2 set. 2025.

SARAVIA, Elvis (organizador). Guia de Engenharia de Prompt (Prompt Engineering Guide) [online]. 2022. Disponível em: <https://www.promptingguide.ai/pt>. Acesso em: 2 set. 2025.

RAGAS. Ragas – stable documentation [online]. Disponível em: <https://docs.ragas.io/en/stable/>. . Acesso em: 2 set. 2025

Observações

(Nenhuma)