

Plano de Ensino

DISCIPLINA: Ciência de dados

CURSO: Processamento de linguagem natural

CARGA HORÁRIA: 80h

CÓDIGO DA DISCIPLINA: COM550

EMENTA

Introdução ao processamento de linguagem natural; aplicações de PLN; gramáticas e análise sintática (classificação de Chomsky); técnicas de análise semântica; modelos de representação de conhecimento; representação lógica; semântica procedimental e representações híbridas.

OBJETIVOS DA DISCIPLINA

Introduzir conceitos de técnicas de interpretação da linguagem humana por sistemas computacionais.

CONTEÚDO PROGRAMÁTICO

1. Introdução ao Processamento de Linguagem Natural
2. *Corpus* e Similaridade Textual
3. Modelos de Linguagem e Representações Vetoriais
4. *Embeddings*
5. Aprendizado de máquina
6. Classificação de textos
7. Análise de sentimentos
8. Revisão

BIBLIOGRAFIA

Bibliografia básica

1. LUGER, G. F. **Inteligência artificial**. São Paulo: Pearson, 2013.

Bibliografia complementar

1. EISENSTEIN, J. **Introduction to Natural Language Processing**. Massachusetts: MIT Press, 2019.

PRÉ-REQUISITOS

Não possui.

CRITÉRIOS DE AVALIAÇÃO

A avaliação da disciplina é formativa e somativa**. Os alunos devem entregar as resoluções de atividades e/ou exercícios no Ambiente Virtual de Aprendizagem semanalmente e realizar, ao final do período letivo, uma prova presencial aplicada nos polos da Univesp.*

**A avaliação formativa ocorre quando há o acompanhamento dos alunos, passo a passo, nas atividades e trabalhos desenvolvidos, de modo a verificar suas facilidades e dificuldades no processo de aprendizagem e, se necessário, adequar alguns aspectos do curso de acordo com as necessidades identificadas.*

***A avaliação somativa é geralmente aplicada no final de um curso ou período letivo. Esse tipo de avaliação busca quantificar o que o aluno aprendeu em relação aos objetivos de aprendizagem do curso. Ou seja, a avaliação somativa quer comprovar se a meta educacional proposta e definida foi alcançada pelo aluno.*

DOCENTE RESPONSÁVEL

Prof. Dr. Marcio Porto Basgalupp

Bacharel em Ciência da Computação pela Universidade Federal de Pelotas (2005), mestre em Ciência da Computação pela PUCRS (2007) e doutor em Ciências de Computação e Matemática Computacional pela Universidade de São Paulo (2010). Realizou doutorado-sanduíche na *University of Kent*, na Inglaterra, em 2009, sob supervisão do Dr. Alex A. Freitas. No ano de 2010, realizou pós-doutorado na NTNU (*Norwegian University of Science and Technology*), na cidade de Trondheim, Noruega. Atualmente, é Prof. Associado III do Instituto de Ciência e Tecnologia da Universidade Federal de São Paulo (Unifesp), campus São José dos Campos, SP. Foi coordenador e um dos idealizadores do Programa de Pós-Graduação em Ciência da Computação (PPG-CC) do ICT-Unifesp desde a sua confecção (e consequente abertura do programa, em 2012) até a aprovação do curso de doutorado, que teve início em 2015. Em 2016, ficou mais um período na *University of Kent*, desta vez como pós-doutorando. Tem coordenado diversos projetos de pesquisa com financiamento das principais agências de fomento do país, em particular CNPq e Fapesp, tanto no âmbito nacional quanto internacional. É um dos pesquisadores principais do Centro de Pesquisa Aplicada (CPA) em Inteligência Artificial, IARA. Seus principais interesses de pesquisa estão relacionados às áreas de Data Science, Aprendizado de Máquina, Computação Bioinspirada e Cidades Inteligentes. Desde março de 2014 é bolsista de produtividade em pesquisa do CNPq.