



UNIVERSIDADE FEDERAL DE SANTA CATARINA – UFSC
CENTRO DE CIÊNCIAS DA EDUCAÇÃO – CED
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO – CIN
CAMPUS UNIVERSITÁRIO - TRINDADE
CEP: 88040-970 - FLORIANÓPOLIS - SC
Fone: (048) 3721-4075 E-mail: cin@contato.ufsc.br

PROGRAMA DE ENSINO

1 IDENTIFICAÇÃO

Disciplina: **CIN8020 – Linguística Computacional**

Carga Horária: 72 h/a – 2 créditos teóricos e 2 créditos práticos

Oferta: Optativa para o Curso Superior de Tecnologia em Ciência de Dados

Pré-requisito: CIN8001 – Programação I

EMENTA

Fundamentos da Linguagem Natural. Morfossintaxe e semântica da linguagem. Interpretação e compreensão textual. Análise de conteúdo. Regras de inferência. Extração e descoberta da informação.

2 OBJETIVOS

2.1 Objetivo Geral

Estudar de maneira aprofundada os conceitos básicos da linguística computacional com vistas a melhorar sua capacidade de descobrir e extrair informações por meio do desenvolvimento de projetos de mineração textual.

2.2 Objetivos Específicos:

2.2.1 Conhecer os fundamentos da linguagem natural.

2.2.2 Estudar os conceitos básicos da linguística computacional.

2.2.3 Implementar projetos de mineração textual que apliquem as técnicas e ferramentas estudadas.

2.2.4 Avaliar os resultados da mineração.

2.2.5 Capacitar-se na execução da mineração de texto com as técnicas e ferramentas aprendidas nos mais diversos cenários de aplicação.

3 CONTEÚDO PROGRAMÁTICO

3.1 Fundamentos da Linguagem Natural

3.1.1 Representações e simbolismos

3.1.2 Aquisição e uso

3.2 Conceitos básicos da Linguística Computacional

- 3.2.1 Fonética e fonologia
- 3.2.2 Léxico, sintaxe e semântica
- 3.2.3 Morfologia, pragmática e sintagmas
- 3.2.4 Estilometria
- 3.2.5 Semiótica
- 3.2.6 Análise de discurso

3.3 Sintaxe e Estrutura da Linguagem

- 3.3.1 Palavras e expressões
- 3.3.2 Orações e sentenças
- 3.3.3 Tipologia de termos
- 3.3.4 Gramática de dependência
- 3.3.5 Gramática de constituição

3.4 Semântica da Linguagem

- 3.4.1 Lemas e lexemas
- 3.4.2 Homônimos, homógrafos e homófonos
- 3.4.3 Heterônimos e heterógrafos
- 3.4.4 Polissemia
- 3.4.5 Capitônimos
- 3.4.6 Sinônimos e antônimos
- 3.4.7 Hipônimos e hiperônimos

3.5 Análise Textual

- 3.5.1 Primórdios
- 3.5.2 Análise textual contemporânea
- 3.5.3 Partes do discurso e etiquetagem
- 3.5.4 Análise morfossintática
- 3.5.5 *Treebanks*
- 3.5.6 *Chunking e parsing*
- 3.5.7 Extração e descoberta da informação
- 3.5.8 Resolução de correferências e anáforas
- 3.5.9 Coerência do discurso

4 BIBLIOGRAFIA

4.1 Bibliografia básica

BENDER, Emily M. **Linguistic fundamentals for natural language processing: 100 essentials from morphology and syntax**. Springer Nature, 2022.

CLARK, Alexander; FOX, Chris; LAPPIN, Shalom (Ed.). **The handbook of computational linguistics and natural language processing**. John Wiley & Sons, 2013.

JURAFSKY, Dan; MARTIN, James H. **Speech and Language Processing**: 3rd ed. draft, 2023. Disponível em: <http://web.stanford.edu/~jurafsky/slp3/>. Acesso em: 20 mar. 2025.

4.2 Bibliografia complementar

DINKELMAN, Daniel. **Computational Linguistics and Artificial Intelligence**: a survey of low-resource language processing, AI ethics in NLP, and neural network transparency regarding language evolution AI. Cineris Multifacet, 2023. 252 p.

ROTHMAN, Denis. **RAG-Driven Generative AI: Build custom retrieval augmented generation pipelines with LlamaIndex, Deep Lake, and Pinecone**. Packt Publishing, 2024.

TAYLOR, Ronald. **Mastering AI with Confidence**: a practical guide to LLMs and natural language processing for developers. Independently published, 2025. 139 p.

THOMAS, Stephen. **Text Mining and Natural Language Processing Resources**. Disponível em: https://github.com/stepthom/text_mining_resources. Acesso em: 20 mar. 2025.

TUNSTALL, Lewis; VON WERRA, Leandro; WOLF, Thomas. **Natural language processing with transformers**. O'Reilly Media, Inc., 2022.