



Universidade Federal de Alagoas  
Instituto de Computação

# Processamento de Linguagem Natural

Prof. Thales Vieira

# Ementa

- Técnicas de pré-processamento: tokenização, limpeza, remoção de stop words, stemização, etc.
- Modelos n-gram e aplicações.
- Medidas de similaridade de palavras e documentos: distância de Levenshtein, matriz termo-documento, bag of words, similaridade do cosseno, TF-IDF.
- Aprendizagem de Máquina para PLN: conceitos básicos, regressão logística, classificador Naïve-Bayes e SVM linear.
- Modelagem de tópicos, classificação e agrupamento de texto: espaços latentes, Alocação Latente de Dirichlet (LDA), análise semântica latente com SVD.
- Redes Neurais e vetores de palavras (word embeddings): word2vec e arquiteturas de redes neurais recorrentes (RNN).
- Reconhecimento de Entidades Nomeadas (NER).
- Etiquetagem gramatical de palavras.
- Outros...

# Bibliografia

- Curso da Intel: <https://software.intel.com/content/www/us/en/develop/training/course-natural-language-processing.html>
- Artigos científicos
- Jurasky, Daniel, and James H. Martin. "Speech and Language Processing: An introduction to natural language Processing." Computational Linguistics and Speech Recognition. Prentice Hall, New Jersey (2000).
- Bird, Steven, Ewan Klein, and Edward Loper. Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.", 2009.
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

# Ferramentas

- Linguagem de programação Python
- Ambiente Google Colab
- Bibliotecas
  - Manipulação de dados: Numpy, Pandas
  - PLN: NLTK, spaCy
  - Machine Learning: Scikit-learn, Keras+TensorFlow

# Avaliação

- Participação individual em aula
- Listas de exercícios individual/equipe (teóricos e práticos via Google Colab)
- Projeto final individual/equipe a ser proposto por cada equipe