

---

# UFMG/DCC/ICEX

## Proposta de disciplina: Processamento de Linguagem Natural

2015/02

---

### Identificação da disciplina

- **Título:** Processamento de Linguagem Natural
- **Nível:** Graduação e Pós-Graduação
- **Professor:** Adriano Alonso Veloso

### Dados gerais da disciplina

- **Objetivo:** O processamento de linguagem natural é o estudo formal sobre as interações entre computadores e as linguagens humanas. O principal objetivo do processamento de linguagem natural é prover aos computadores a capacidade de derivar significado ou semântica a partir de texto ou áudio em linguagem natural. Assim, neste curso serão cobertos os elementos e aspectos fundamentais do Processamento de Linguagem Natural, bem como sua aplicação na resolução de problemas da Ciência da Computação.
- **Carga Horária:** 60 créditos
- **Livros texto:** Teórico [1,2] e aplicado [3]
- **Pré-requisitos:** Inglês. Álgebra Linear básica. Conhecimento prévio de algoritmos e programação (AEDs 3).
- **Forma de avaliação:** Duas provas (30 pontos cada), e um projeto de complexidade desafiadora (40 pontos)
- **Horário preferido:** terças e quintas as 17 horas.

### Ementa da disciplina

1. Introdução: processamento básico de texto, expressões regulares, normalização e radicalização, segmentação de sentenças, distância de edição, alinhamentos.
2. Modelos de linguagem: N-grams, formas de avaliação e perplexidade, interpolação, correção de ortografia.
3. Classificação de texto: Naive Bayes, formas de avaliação, análise de sentimentos em texto e áudio.
4. Extração de informação: reconhecimento de entidades nomeadas, formas de avaliação, extração de relação.

5. Discurso: partes do discurso, *parsing* e estrutura sintática, *parsing* probabilístico, formas de avaliação.
6. Recuperação de informação: matrizes termo-documento, índice invertido, *ranking*, ponderação, modelo espaço-vetorial, formas de avaliação.
7. Semântica: sentido, desambiguação, representações vetoriais semânticas, relações, thesaurus, wordnet, similaridade distribuída, *question answering*, formulação de perguntas, recuperação de passagens e extração de respostas, sumarização.

## **Referências**

**[1] Speech and Language Processing (2nd Edition): Daniel Jurafsky, James Martin.**

**[2] Foundations of Statistical Natural Language Processing: Christopher Manning, Hinrich Shuetz.**

**[3] Natural Language Processing with Python: Steven Bird, Ewan Klein, Edward Loper.**