

Disciplina:

Processamento de linguagem natural

Professor: Gabriel Assunção



Quem sou eu?

Gabriel Oliveira Assunção

Graduação em Estatística UFMG;

Mestrado em Estatística UFMG;

Doutorando em Estatística UFMG

Cientista de Dados na Localiza&Co



[/gabrieloassuncao](https://www.linkedin.com/in/gabrieloassuncao)

Apresentação do curso

Objetivo

- Expressões regulares.
- Medidas de similaridade textual.
- Parsing, tokenização, lematização, stemming.
- Reconhecimento de entidades nomeadas.
- Modelos de Linguagem.
- NLP usando redes neurais.
- Análise de sentimento.

Apresentação do curso

Módulos

1. Introdução
2. Expressão regular
3. Processamento de texto (Lista 1: 10 pontos)
4. Representação textual (Lista 2: 10 pontos)
5. Classificação de texto
6. Modelos de NLP (Lista 3: 10 pontos)

Apresentação do curso

Seminário

- a. Objetivo: apresentar alguma aplicação de NLP.
- b. Duas últimas aulas.
- c. Apresentação do tema (20~25 minutos)
- d. Debate (10~15 minutos)
- e. Avaliação:
 - i. 20 pontos apresentação (material + apresentação em sala)
 - ii. 10 pontos relatório das outras apresentações.

Apresentação do curso

Aplicação a dado

- a. Objetivo: realizar uma aplicação do conteúdo da disciplina em alguma base de dados.
- b. Avaliação:
 - i. 10 pontos apresentação do objetivo do trabalho. Banco de dados + análise a ser feita.
 - ii. 20 pontos: código utilizado para a análise.
 - iii. 10 pontos: relatório com as análises.

Sobre vocês

Ferramentas da disciplina

Instalar o [Python](#);

Instalar o [Anaconda](#);

Instalar o [Pip](#);

Instalar [Jupyter](#);

O que é NLP?

O que é NLP?

NLP: Natural Language Process (Processamento de linguagem natural)

NLP é o ramo da ciência da computação que tem como objetivo dar habilidade da máquina a compreender a linguagem humana.

Objetivo: Ter computadores que entendem a linguagem natural para executar tarefas úteis. Exemplo: traduzir textos, responder a perguntas, fazer agendamentos, realizar compras, receber comandos.

O que é NLP?

Linguagem artificial

Java

Python

C

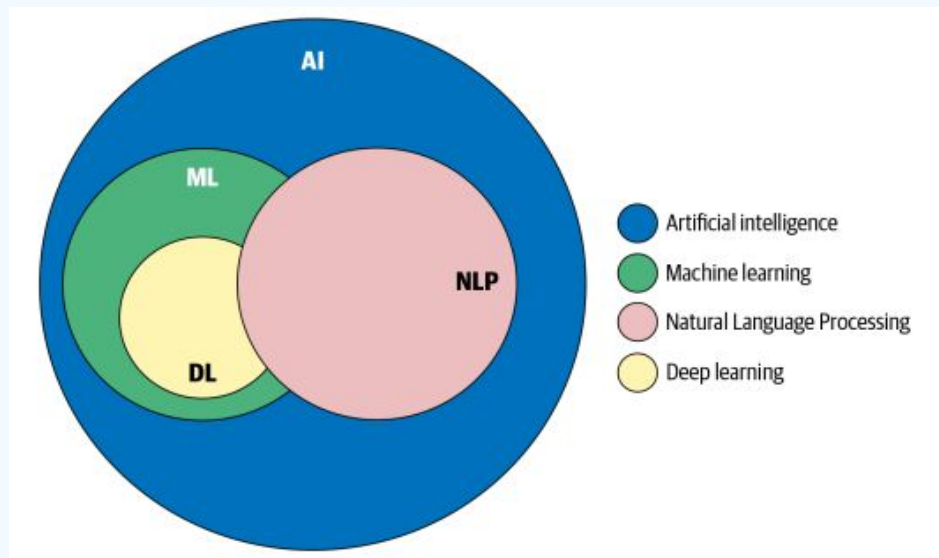
R

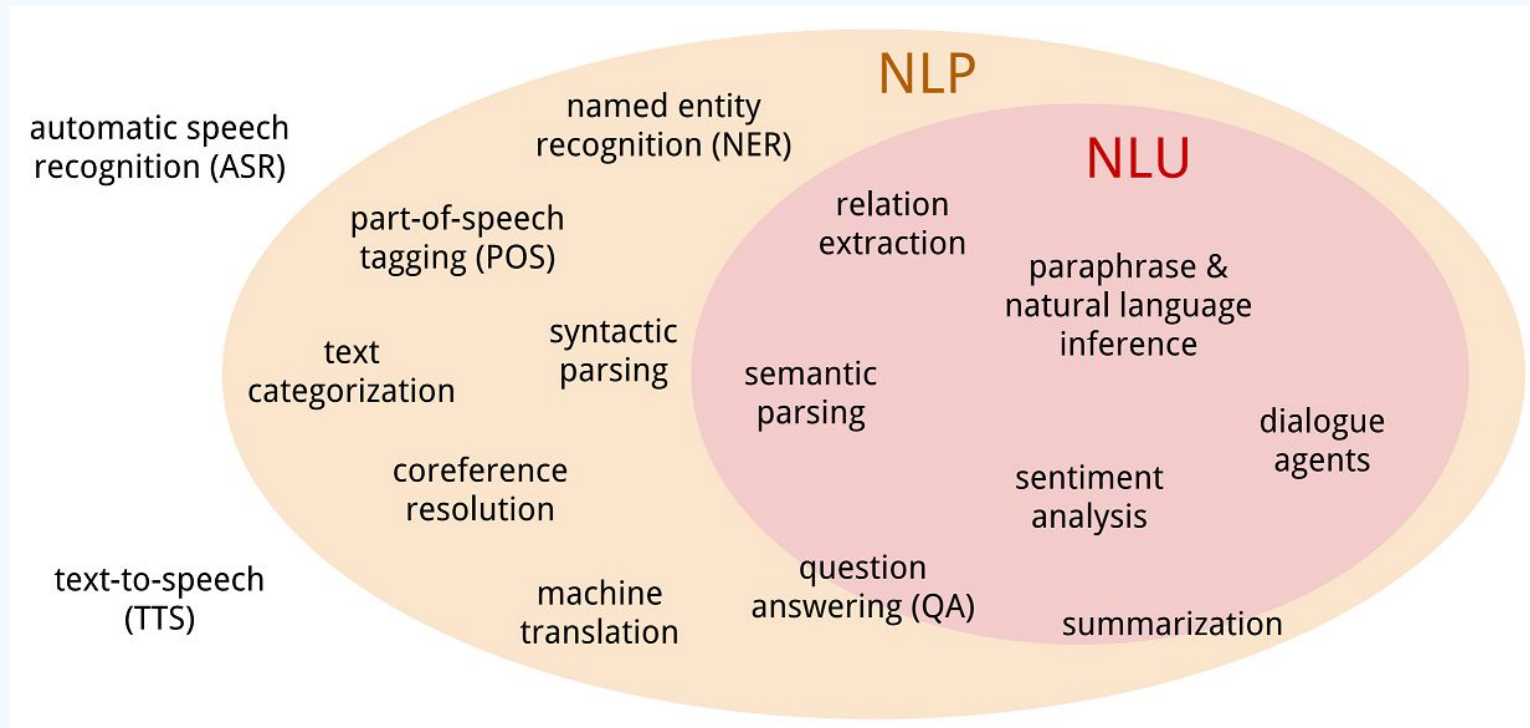
código binário.

Linguagem natural

Textos

Falas





Por que a dificuldade para entender textos?

- Os computadores lidam principalmente com dados estruturados (tabelas, bancos de dados).
- NLP - dados não estruturados
 - Postagens de mídia social
 - Artigos de notícias
 - e-mails
- Aprender a estrutura e a gramática da linguagem.

Por que a dificuldade para entender textos?

- Trocadilhos
- Sarcasmo
- Metáforas
- Variedade linguística (regionalismo)
- Erros ortografia
- Ambiguidade
- Estrangeirismo
- Linguagem culta e informal
- Gírias

Problemas que o NLP pode resolver

- Análise de sentimentos
- Tradução
- Perguntas e resposta (Q&A)
- Sumarização de texto
- Classificação do texto