
Processamento de Linguagem Natural

Aula 01
Introdução

Agenda

- Apresentação
- Disciplina
- O que é Processamento de Linguagem Natural? A natureza da linguagem humana
- Dificuldade em entender a linguagem

Apresentação da Matéria

Apresentação

Eu sou a **Keyla Macharet**

Sou mestre em Sistemas de Computação pelo Programa de Pós-Graduação em Engenharia Elétrica da UFMG e graduada em Ciência da Computação pelo Departamento de Ciência da Computação da UFMG.

Estou na **Take** há mais de 16 anos

Team Leader e Chatbot Developer



Disciplina

	Seg	Ter	Qua	Qui	Sex	Sab	Dom
Agosto				1	2	3	4
	5	6	7	8	9	10	11
	12	13	14	15	16	17	18
	19	20	21	22	23	24	25
	26	27	28	29	30	31	01
Setembro	02	03	04	05	06	07	08
	09	10	11	12	13	14	15

Objetivos

- Abordar conceitos de processamento de linguagem natural e suas aplicações práticas.
- Habilitar o aluno a fazer o tratamento de textos e extrair informações relevantes.
- Demonstração de software, práticas simuladas, apresentação teórica.

Disciplina

- Aula 1 (01/08)
 - Apresentação da Disciplina
 - Algoritmos e técnicas de processamento de linguagem natural.
 - Expressões regulares.
- Aula 2 (08/08)
 - Normalização e Pré Processamento de Texto
 - Conceitos de Parsing, tokenização, lematização e stemming.
- Aula 3 (22/08)
 - Medidas de similaridade textual
 - Similaridade Léxica e Similaridade Semântica



Disciplina

- Aula 4 (29/08)
 - Marcação textual - Part of Speech Tagging
 - Unigram Tagger e N-Gram Tagger
- Aula 5 (05/09)
 - Extração de informação e Reconhecimento de entidades nomeadas.
- Aula 6 (12/09)
 - Arquitetura de aplicação para processamento de Linguagem Natural.
 - Ferramentas comerciais
 - Estado da arte em NLP

Teremos sempre a primeira parte da aula com conteúdo teórico sobre o assunto proposto e a segunda metade com parte prática do que foi aprendido.



Fontes e Bibliografia

- **Natural Language Processing with Python** – Analyzing Text with the Natural Language Toolkit (Steven Bird, Ewan Klein, and Edward Loper) - <https://www.nltk.org/book/>
- **Speech and Language Processing - An Introduction to Natural Language Processing**, Computational Linguistics, and Speech Recognition (Daniel Jurafsky e James H. Martin) - <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>
- **Natural Language Processing** - CS224N - Stanford University - <https://see.stanford.edu/Course/CS224N>



Material

Vamos utilizar o Google Classroom para centralizar os materiais e as entregas de tarefas das aulas

<https://classroom.google.com>

hyy3v53

Avaliação

- Participação nas aulas (20 pts);
- Desenvolvimento dos exercícios das aulas de NLP (no laboratório) (50 pts);
- Trabalho final em grupo (parte no laboratório) (30 pts)

Avaliação - Trabalho Final

- Cada aluno deve pesquisar e desenvolver uma aplicação prática que utilize conceitos e tarefas de NLP aprendidas durante as aulas.
- O trabalho pode ser individual ou em dupla.
- Ele deve ser feito em Python, e de forma que após a entrega eu consiga executá-lo.
- No último dia de aula, cada aluno (ou dupla) deverá apresentar para a turma por no máximo 5 minutos o que é a aplicação que está desenvolvendo.
- Devem ser entregues:
 - Código e todos os insumos que eu preciso para executar a aplicação
 - Um docs de 1 página explicando para que serve a sua aplicação, e como você desenvolveu ela (conceitos, bibliotecas que usou, etc.)



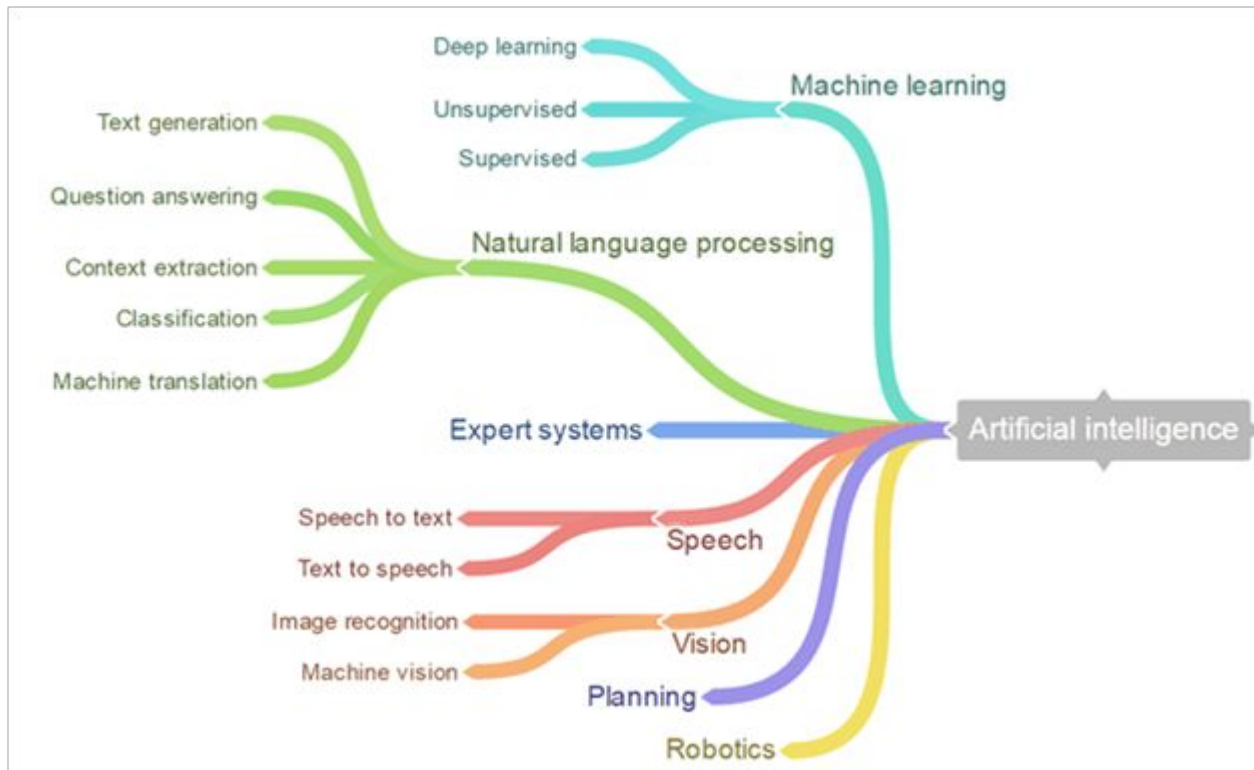
Linguagem de Programação

- Usaremos Python
 - Em aulas vamos ter a parte prática (codificar em sala).
- Instale o Anaconda + Jupyter Notebook no seu computador
 - <https://www.anaconda.com/distribution/>
- Ou utilize o Google Colab
 - <https://colab.research.google.com>



Processamento de Linguagem Natural

O que é Processamento de Linguagem Natural?



Objetivo da PLN

Linguagem artificial

Java, Python, C, JavaScript, R, VBA, código binário.

Linguagem natural

Linguagem falada/escrita por pessoas (humanas).

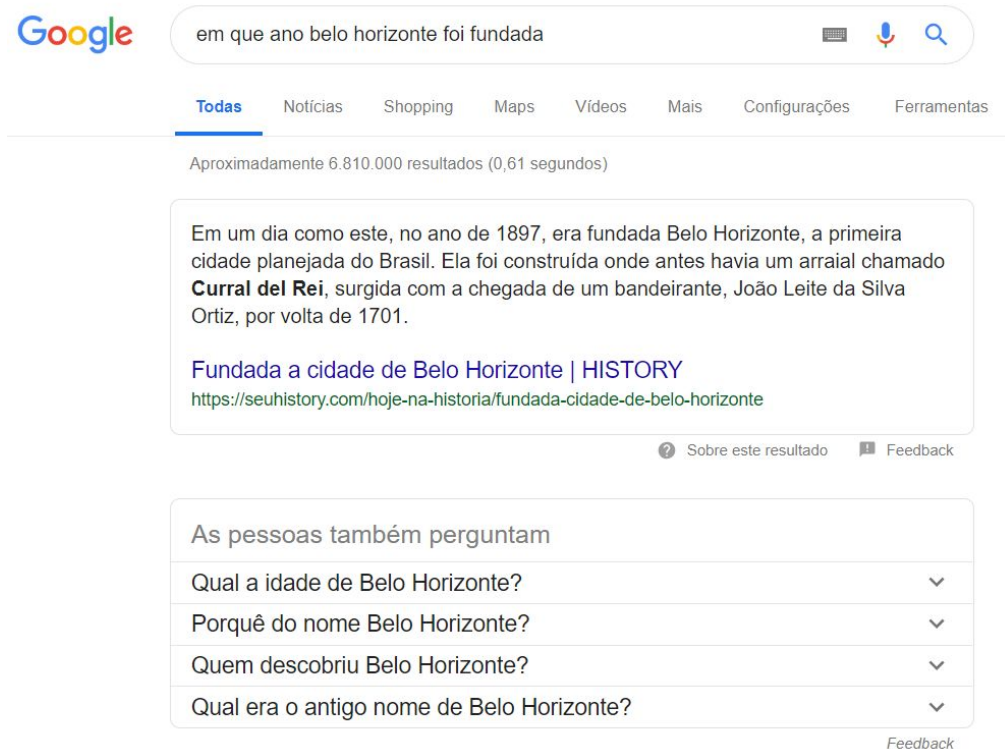
Construir mecanismos artificiais que permitam o entendimento de linguagem natural para realizar tarefas ou aplicações “próximas” ao entendimento humano.

O que é Processamento de Linguagem Natural?

- Objetivo é permitir que o computador “entenda” linguagem natural e realize tarefas úteis
 - Fazer agendamentos, comprar coisas
 - Traduzir textos
 - Responder perguntas (Siri, Cortana, Google Assistant...)

Entender **completamente** o **sentido** da linguagem é um objetivo difícil

Exemplo de sistemas utilizando PLN



The image shows a Google search interface. The search bar contains the text "em que ano belo horizonte foi fundada". Below the search bar, there are tabs for "Todas", "Notícias", "Shopping", "Maps", "Vídeos", "Mais", "Configurações", and "Ferramentas". The "Todas" tab is selected. Below the tabs, it says "Aproximadamente 6.810.000 resultados (0,61 segundos)". The main result is a snippet from "Fundada a cidade de Belo Horizonte | HISTORY" with the URL "https://seuhistory.com/hoje-na-historia/fundada-cidade-de-belo-horizonte". Below the snippet, there are links for "Sobre este resultado" and "Feedback". At the bottom, there is a section titled "As pessoas também perguntam" with four related questions, each with a dropdown arrow.

Google

em que ano belo horizonte foi fundada

Todas Notícias Shopping Maps Vídeos Mais Configurações Ferramentas

Aproximadamente 6.810.000 resultados (0,61 segundos)

Em um dia como este, no ano de 1897, era fundada Belo Horizonte, a primeira cidade planejada do Brasil. Ela foi construída onde antes havia um arraial chamado **Curral del Rei**, surgida com a chegada de um bandeirante, João Leite da Silva Ortiz, por volta de 1701.

Fundada a cidade de Belo Horizonte | HISTORY
<https://seuhistory.com/hoje-na-historia/fundada-cidade-de-belo-horizonte>

Sobre este resultado Feedback

As pessoas também perguntam

- Qual a idade de Belo Horizonte?
- Porquê do nome Belo Horizonte?
- Quem descobriu Belo Horizonte?
- Qual era o antigo nome de Belo Horizonte?

Feedback

Exemplo de sistemas utilizando PLN



What is NLP



[Browse Examples](#) [Surprise Me](#)

Assuming "NLP" is an acronym | Use as a [financial entity](#) instead

Input interpretation:

NLP (acronym)

Possible interpretations:

neuro-linguistic programming

natural language processing

[Download Page](#)

POWERED BY THE WOLFRAM LANGUAGE

Related Queries:

= Morse code NLP

= 36^NLP

= n^i*p

= CAPTCHA NLP

= tell me a chemistry joke!



do you like me?



[Browse Examples](#) [Surprise Me](#)

Assuming "do you like me" is a phrase | Use as a [question about Alpha](#) instead

Assuming Do you like me? | Use [Do you like people \(humans, children, ...\)?](#) instead

Input interpretation:

Do you like me?

Response:

Of course; I like all humans who ask me questions I can answer.

[Download Page](#)

POWERED BY THE WOLFRAM LANGUAGE

Related Queries:

= Do you like people (humans, children, ...)?

= random Pokemon curve

= Do you like (enjoy, ...)?

= is nim chimpsky alive still?

= Egyptian God cards



PUC Minas

Tipos de Buscas Realizadas

- Busca usando termos ou palavras-chave:
 - Relativamente fácil para as plataformas atuais.
 - Os sistemas clássicos usaram termos simples.
- Busca usando perguntas completas:
 - Relativamente difícil.
 - São necessários resolução de inferências, síntese e resumo de informações de diferentes fontes.
- Exemplos de Ferramentas
 - Siri, Google Assistant, Cortana

O que é Processamento de Linguagem Natural?

- NLP (Natural Language Processing) é uma área de estudo que está na interseção de 3 áreas
 - Ciência da Computação
 - Inteligência Artificial (Machine Learning, Deep Learning)
 - Linguística

O que é Processamento de Linguagem Natural?

- PLN é comumente associado com:
 - Linguística computacional;
 - Tecnologia da linguagem;
 - Engenharia da linguagem.
- Linguagem é comumente é usada em contraste com:
 - Fala (Speech);
 - Tecnologia da fala e da linguagem.

Na nossa matéria trataremos somente o processamento de
linguagem textual

O que é Processamento de Linguagem Natural?

- Objetivo de pesquisa:
 - Entender como opera a linguagem humana (escrita ou falada).
- Objetivo de desenvolvimento (engenharia):
 - Construção de sistemas que analisem/gerem linguagem;
 - Reduzir a brecha homem-máquina.



Quais os desafios do Processamento de Linguagem Natural?

- A linguagem humana é difícil de entender.
- A linguagem é aprendida intuitivamente
 - Fácil para crianças, difícil para computadores
- Sarcasmo
 - “Alguns causam felicidade aonde quer que vão. Outros causam sempre que se vão”
~Oscar Wilde
- Trocadilhos
 - “Tudo na vida muda, até a bermuda!”

Quais os desafios do Processamento de Linguagem Natural?

Temos desafios extras no Português, principalmente pela falta de bibliotecas para a nossa língua

Quais os desafios do Processamento de Linguagem Natural?

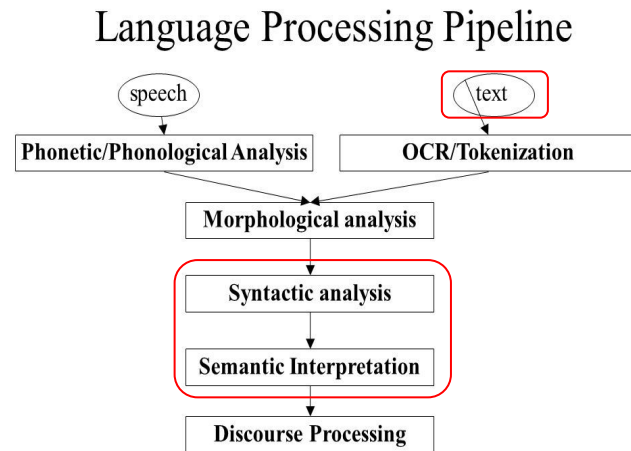
- Assim como a Visão Computacional, entendimento perfeito da linguagem é um problema conhecido como “*AI-complete*” ou “*AI-hard*” (analogia com problemas NP-completo ou NP-difíceis).
- *AI-Complete*
 - São desenvolvidos de forma a incluir **Visão Computacional** e **Natural Language Understanding**.
 - Trabalham com circunstâncias inesperadas enquanto resolvem qualquer problema do mundo real
 - Não podem atualmente ser resolvidos somente com tecnologia, ainda necessitando de computação humana (quando certos passos do processamento dependem de ação humana)



Níveis de PLN

A pesquisa em Pln está voltada, essencialmente, a três aspectos da comunicação em língua natural

- Som: fonologia
- Estrutura: morfologia e sintaxe
- Significado: semântica e pragmática



Níveis de PLN

- A Fonologia está relacionada ao reconhecimento dos sons que compõem as palavras de uma língua.
- A morfologia reconhece as palavras em termos das unidades primitivas que a compõem (caçou → caç+ou).
- A sintaxe define a estrutura de uma frase, com base na forma como as palavras se relacionam nessa frase (Sujeito, predicado).
- A semântica associa significado a uma estrutura sintática, em termos dos significados das palavras que a compõem (“um animal perseguiu/capturou outro animal”).
- A pragmática verifica se o significado associado a uma estrutura sintática é realmente o significado mais apropriado no contexto considerado (no contexto predador-presa, “perseguiu/capturou” → “comeu”).

