



VEJA OS PLANOS

PROGRAMAÇÃO _

DATA SCIENCE _

DEVOPS _

MOBILE _

FRONT-END _

INTELIGÊNCIA ARTIFICIAL _

UX & DESIGN _

INOVAÇÃO & GESTÃO _

Artigos > **Data Science**

PLN: o que é Processamento de Linguagem Natural?



goncalves-thiago

Atualizado em 17/08/2023

COMPARTILHE



Esse artigo faz parte da
Formação Data Science

Introdução

Neste artigo você vai descobrir **o que é PLN, a origem e os desafios**.

O [Processamento de Linguagem Natural](#) (PLN ou NLP) mescla **ciência da computação, inteligência artificial e linguística**, se dedicando à geração e compreensão automática da linguagem natural. Mas e no dia a dia, onde encontramos o PLN?

Confira neste artigo:

- [Introdução](#)
- [O que é PLN](#)
- [Quais os objetivos do PLN](#)
- [Como o PLN funciona?](#)
- [Quais os principais usos do PLN?](#)
- [Aplicação do PLN no dia a dia!](#)
- [Origem](#)
- [Evolução dos modelos](#)
- [Qual a importância do PLN e benefícios para as empresas?](#)
- [Quais os desafios?](#)
- [Conclusão](#)

Fonte: Aliz.ia



Quiz

JORNADA TECH

Teste seus conhecimentos em tecnologia
E prepare-se para o próximo nível da sua carreira

- Programação
- Front-end
- Inteligência Artificial
- Data Science
- Inovação & Gestão

Comece agora

alura

O que é PLN

Antes de mais nada, é importante ter um panorama a respeito do assunto e entender o que significa PLN, para então compreender sua aplicação.

Segundo [estudos](#), o PLN (Processamento de Linguagem Natural) é uma área da Ciência da Computação que estuda o desenvolvimento de programas computacionais que analisam, reconhecem e/ou geram textos em linguagens humanas, ou naturais.

Vale lembrar que a ambiguidade da linguagem natural torna o PLN diferente do processamento das linguagens de programação, definidas justamente para evitar qualquer ambiguidade. A língua muda com o passar do tempo e varia de forma regional. Por isso, o PLN desafia as máquinas a lidar com a ambiguidade, a gramática, o contexto e a riqueza semântica da linguagem

humana, sendo uma área de pesquisa em constante evolução. À medida que a tecnologia avança, o PLN desempenha um papel cada vez mais importante em melhorar a interação entre humanos e máquinas.

Quais os objetivos do PLN

O mesmo estudo citado no tópico anterior aponta que alguns dos principais objetivos em PLN são: a recuperação de informações a partir de textos, a tradução automática, a interpretação de textos e a realização de inferências a partir de textos. Uma outra forma de enxergar os objetivos é através de tópicos macro, tais como:

- Entendimento da linguagem humana;
- Geração de linguagem humana;
- Processamento de linguagem humana.

É essencial destacar que, do ponto de vista linguístico, o foco das pesquisas em PLN podem variar em diferentes níveis de análise, podendo ser: fonético ou fonológico, morfológico, sintático, semântico ou pragmático. Dessa forma, as aplicações de PLN podem variar conforme a necessidade, utilizando níveis específicos de análise para reduzir a complexidade de tratamento.

Como o PLN funciona?

Do ponto de vista computacional podemos dividir o PLN em algumas etapas:

- No **pré-processamento de texto** são feitas preparações como a remoção de pontuações, transformação das maiúsculas em minúsculas para a padronização e remoção de palavras desnecessárias, como artigos e preposições, também chamados de stop words.
- Numa próxima etapa é feito o processo de **tokenização** no qual o texto é dividido em unidades menores, chamadas “tokens”, que podem ser palavras, frases ou subpalavras. Isso facilita a análise e a compreensão do texto.
- Posteriormente podem ser feitas **análises morfológicas, sintáticas ou semânticas**. Também pode ser aplicado o Machine Learning, com o intuito de fazer uma predição com base nos dados de texto.

Quais os 3 níveis de reconhecimento na técnica de processamento de linguagem natural?

Após o processamento inicial podem ser feitas análises no conjunto de dados. As análises praticadas no PLN estão relacionadas ao que chamamos de níveis de reconhecimento. Os níveis de reconhecimento e análises são três: Nível morfológico, sintático e o semântico.

- Na análise morfológica as palavras são examinadas para identificar sua forma básica (lematização) e suas características gramaticais como a parte do discurso, gênero, número, tempo verbal, etc.
- Na análise sintática a estrutura gramatical é utilizada para entender o relacionamento entre as palavras e a organização das frases.
- Já na análise semântica a ideia é tentar compreender o significado das palavras e frases, além do contexto em que elas são utilizadas.

Um ponto importante é a ambiguidade como característica comum na linguagem natural. Por isso, temos uma etapa de **desambiguação**, na qual o objetivo é tentar resolver as ambiguidades e determinar o significado mais apropriado com base no contexto. Assim, podemos realizar a extração de informações específicas, como nomes de pessoas, datas, localizações, eventos importantes, etc.

Após as análises, muitos sistemas de PLN utilizam Machine Learning. A ideia é ajustar um modelo matemático aos dados para realizar tarefas preditivas. A classificação de texto, análise de sentimentos e a tradução são exemplos da aplicação de Machine Learning.

Quais os principais usos do PLN?

Você já deve ter conversado com um chatbot para obter atendimento de uma empresa. Até mesmo o ChatGPT e outros modelos de linguagem mais

modernos, são aplicações de PLN em conjunto com Machine Learning. Outro exemplo são os assistentes virtuais como a Alexa e o Google Assistant. Nesse caso, existe uma combinação de transcrição de áudio e interpretação de texto para concluir uma tarefa específica.

Esses temas parecem complexos, mas a ideia principal é utilizar o PLN como uma ferramenta para diminuir a complexidade de tarefas. Uma possibilidade para quem tem que ler muitos textos é utilizar ferramentas criadas com PLN para resumir textos, realizar correções gramaticais ou mesmo categorizar e classificar textos em tópicos ou temas específicos. No caso das mídias sociais, podemos acompanhar e analisar as conversas em redes sociais. Assim, empresas e marcas podem compreender melhor a percepção pública sobre elas.

Aplicação do PLN no dia a dia!

Quantas mensagens de texto você leu hoje? Quantos áudios foram enviados por apps de mensagens? Quantos e-mails foram recebidos? Quantos posts em redes sociais foram enviados?

Para todas as perguntas acima, a provável resposta é que foram muitos. O [webfx](#) mostra em tempo real o quanto de dados está sendo trafegado na internet a cada instante, se você entrar, repare que é um volume gigantesco, e grande parte desses dados são textos, áudios ou vídeos.

A disponibilidade de dados em linguagem natural nunca foi tão grande, por isso, tratar e desenvolver algoritmos que possibilitem o desenvolvimento de aplicações úteis em nosso dia a dia com toda essa informação, é parte fundamental do processamento de linguagem natural.

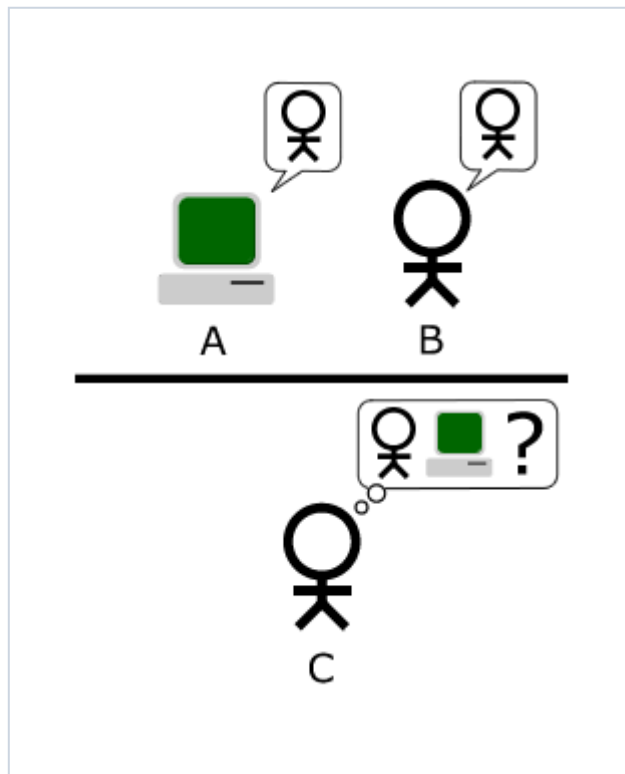
Mas uma dúvida que fica é: de onde veio esta ideia de misturar computação com linguística?

Origem

Em 1950 [Alan Turing](#) publicou o artigo [Computing Machinery and Intelligence](#), sendo [pioneiro em inteligência artificial](#), trazendo a proposta de um teste conhecido atualmente como [Teste de Turing](#).

O desafio que consiste na identificação de uma máquina ou uma pessoa baseada apenas nas perguntas e respostas de uma conversa homem-

máquina, deu origem ao processamento de linguagem natural como conhecemos hoje.



Fonte: Wikipédia

Surge, então, nos anos 60 os primeiros sistemas bem sucedidos, dentre eles o que se destaca é a [ELIZA](#) que simula uma psicoterapeuta.

Usando informações sobre o pensamento e emoções humanas, ELIZA criava interações surpreendentemente humanas. Quando o usuário ou usuária extrapolava a base de conhecimento do programa, ELIZA fornecia respostas genéricas para manter a conversa ainda mais humana.

Até os anos 80, grande parte dos sistemas de PLN eram baseados em conjuntos complexos de regras, e no final da década, os algoritmos de **machine learning** deram início a uma nova e promissora fase na área. Nesta fase, os algoritmos eram relativamente simples, como árvore de decisão, e já produziam sistemas de regras semelhantes ou até melhor que as regras escritas “à mão” no início da década.

Evolução dos modelos

Desde então, modelos estatísticos, técnicas de [machine learning](#) e mais recentemente **deep learning** dominam a área, criando sistemas cada vez mais próximos de passar no teste de Turing.

Os novos modelos nos impressionam com tradutores cada vez melhores, **chatbots** humanizados e classificadores capazes de perceber sentimentos em relação a um produto analisando suas avaliações textuais.

Um caso muito interessante ocorreu em Setembro de 2019, quando o Allen Instituto Para Inteligência Artificial (AI2) revelou ao mundo o programa chamado [Aristo](#). Ele pode resolver corretamente um teste científico com taxa de acerto maior que 90%, embora o teste seja de um 9º ano escolar, no Brasil, e pareça relativamente simples para um adulto, essa tarefa é extremamente difícil para um computador.

Para encontrar as respostas corretas a cada uma das perguntas, foi necessário analisar bilhões e bilhões de documentos. Retirar informações úteis desses bilhões de documentos só foi possível com a intersecção de áreas como computação, linguística e inteligência computacional.

O PLN é justamente essa área interdisciplinar do conhecimento que trabalha com a interação entre computadores e humanos usando a linguagem natural. A partir de modelos linguísticos como a teoria da gramática generativa de [Noam Chomsky](#) e modelos estatísticos, o PLN processa, analisa e gera valor tendo como input dados em linguagem natural, sejam textos ou áudios.

Qual a importância do PLN e benefícios para as empresas?

O PLN pode ser usado para acelerar e automatizar diversos processos numa empresa. Muitas informações importantes estão em formato de texto em e-mails, relatórios, postagens em redes sociais e comentários de clientes. O PLN permite extrair insights valiosos a partir desses documentos.

Chatbots também são amplamente adotados por empresas de diversos tamanhos. O PLN pode torná-los ainda mais direcionados e inteligentes. Mas não para por aí, essas ferramentas permitem que empresas ofereçam suporte 24 horas por dia por um baixo custo. Além disso, o PLN permite a personalização de conteúdo através da análise das preferências de clientes, permitindo a construção de recomendadores de conteúdo.

De forma resumida, o PLN oferece benefícios que podem impulsionar o crescimento e o sucesso das empresas, melhorando a experiência do cliente, otimizando processos internos e facilitando a análise de dados textuais.

Quais os desafios?

O processamento de linguagem natural não é um problema solucionado, principalmente quando tratamos de **análise semântica**. No contexto da linguística, a semântica é o estudo do significado usado por seres humanos para se expressar através da linguagem. Para compreender ironia ou sarcasmo, por exemplo, não são analisadas apenas as palavras utilizadas na frase, mas o contexto em que elas estão aplicadas.

A generalização dos **modelos de NLP** também é um problema, embora avance a passos largos devido às técnicas mais avançadas de Deep Learning lançadas nos últimos anos. Criar modelos treinados em uma linguagem, que sejam genéricos o suficiente para serem aplicados de forma satisfatória em outros idiomas, ainda é um problema desafiador e importante, visto que nem todas as línguas têm base de dados boas o suficiente para treinar os modelos de deep learning mais avançados.

A especificidade de cada língua dificulta o trabalho com linguagem natural, o que o torna cheio de desafios que após serem superados, abrirão caminho para o desenvolvimento de muitas aplicações com potencial de transformar a maneira como lidamos com tecnologia.

Conclusão

O processamento de linguagem natural é esta área que mescla computação, IA e linguística, cheia de desafios, com aplicações que fazem parte do nosso dia a dia e um mundo de possibilidades com potencial de transformar o modo como usamos celular, computador e outras tecnologias. Gostou deste artigo e quer conhecer mais sobre PLN? Aqui na **Alura** temos muito conteúdo para você:

- [Corretor ortográfico com nlp](#)
- [Modelos de linguagem](#)
- [Análise de sentimento](#)

Para ter acesso a conteúdos interessantes sobre [ciência de dados](#), me siga no Twitter @tgcsantos, sempre tem conteúdo interessante rolando por lá.



[Artigo Anterior](#)

SQL JOIN: Aprenda INNER, LEFT, RIGHT, FULL e CROSS

[Próximo Artigo](#)

Transações no SQL: Mantendo os dados íntegros e consistentes

Leia também:

[O que é IA Generativa? A importância e o uso das Inteligências Artificiais como](#)

- [ChatGPT, MidJourney e outras](#)
- [Por onde começar os estudos na área de dados?](#)
- [O que é Ciências de Dados?](#)
- [Qual a diferença entre Data Science, Machine Learning e Inteligência Artificial?](#)
- [Direto ao ponto: o que é Machine Learning com exemplos reais](#)

Veja outros artigos sobre
[Data Science](#)

Quer mergulhar em tecnologia e aprendizagem?

Receba a newsletter que o nosso CEO escreve pessoalmente, com insights do mercado de trabalho, ciência e desenvolvimento de software

Escreva seu email

ME INSCREVA

Nossas redes e apps



Institucional

Sobre nós

Trabalhe conosco

Para Empresas

A Alura

Formações

Como Funciona

Todos os cursos

[Para Sua Escola](#)

[Política de Privacidade](#)

[Compromisso de Integridade](#)

[Termos de Uso](#)

[Documentos Institucionais](#)

[Status](#)

[Depoimentos](#)

[Instrutores\(as\)](#)

[Dev em <T>](#)

[Luri, a inteligência artificial da Alura](#)

Conteúdos

[Alura Cases](#)

[Imersões](#)

[Artigos](#)

[Podcasts](#)

[Artigos de educação corporativa](#)

Fale Conosco

[Email e telefone](#)

[Perguntas frequentes](#)

Novidades e Lançamentos

Email*

ENVIAR

CURSOS

Cursos de Programação

[Lógica](#) | [Python](#) | [PHP](#) | [Java](#) | [.NET](#) | [Node JS](#) | [C](#) | [Computação](#) | [Jogos](#) | [IoT](#)

Cursos de Front-end

[HTML, CSS](#) | [React](#) | [Angular](#) | [JavaScript](#) | [jQuery](#)

Cursos de Data Science

Ciência de dados | BI | SQL e Banco de Dados | Excel | Machine Learning | NoSQL | Estatística

Cursos de Inteligência Artificial

IA para Programação | IA para Dados

Cursos de DevOps

AWS | Azure | Docker | Segurança | IaC | Linux

Cursos de UX & Design

Usabilidade e UX | Vídeo e Motion | 3D

Cursos de Mobile

React Native | Flutter | iOS e Swift | Android, Kotlin | Jogos

Cursos de Inovação & Gestão

Métodos Ágeis | Softskills | Liderança e Gestão | Startups | Vendas

CURSOS UNIVERSITÁRIOS FIAP

Graduação | Pós-graduação | MBA