

# IF704 - Processamento de Linguagem Natural

Wilton Ramos da Silva

7 de maio de 2018

## 1 Introdução

Processamento de Linguagem Natural (doravante PLN) é uma cadeira eletiva do curso de graduação em Ciência da Computação e se enquadra como uma subárea da Inteligência Artificial, a qual possibilita que máquinas e programas consigam processar e reproduzir informações linguísticas de linguagens não artificiais. Dessa forma, a disciplina trata não somente de questões relacionadas à Computação, mas apresenta uma interdisciplinaridade com a área de Linguística. Destarte, PLN explicita tópicos formais das línguas naturais, como sintaxe e semântica, e tópicos funcionais, como o discernimento de discursos, atrelando-os a assuntos de Estatística, álgebra e programação.

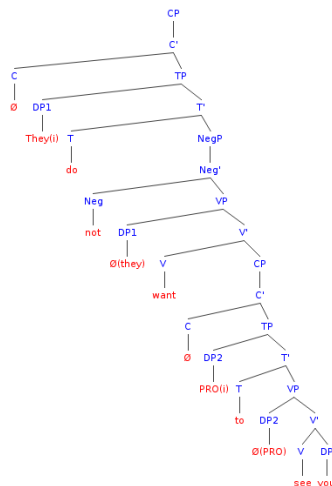


Figura 1: Exemplo de árvore sintática

## 2 Relevância

O principal ponto que faz com que PLN integre o currículo do curso de Ciência da Computação está relacionado com a formação específica em Inteligência Artificial, visto que aquela é uma subárea de extrema importância desta por lidar com tarefas que envolvam a forma como nós nos comunicamos, que é, majoritariamente, por meio de linguagem natural. Outros dois motivos, não menos importantes, que torna essa disciplina relevante para a formação de cientistas da computação é a sua ampla aplicação e o crescente interesse do mercado sobre. Entre as aplicações, podemos exemplificá-las com algumas que seguem:

- programas de tradução;
- análise de sentimento;
- sumarização automática;
- filtro de SPAM;
- reconhecimento de voz.

Além disso, o mercado para essa área está em contínuo crescimento (é esperado que movimente cerca de \$22 bilhões em 2025<sup>1</sup>) e desperta o interesse de empresas como Google, Facebook e Microsoft.

### 3 Relação com outras disciplinas

Tabela 1: Relação de PLN com outras disciplinas

Cadeiras	Relação
ET586 - Estatística e probabilidade para computação	Um dos conhecimentos prévios necessários para conseguir desenvolver programas de PLN está relacionado com noções básicas de probabilidade, uma vez que é necessário métodos para desambiguar sentenças, por exemplo.
IF669 - Introdução à programação	Como a cadeira visa, inclusive, o desenvolvimento de programas que processem linguagem natural, é imprescindível conhecimento de programação.
IF759 - Processamento de voz	PLN funciona como um dos pré-requisitos para essa cadeira, já que aquela trata de temas basilares à esta, como processamento sintático e interpretação semântica.
MA026 - Cálculo diferencial e integral I	Noções como derivadas e como máximos e mínimos de funções são aplicadas nos estudos de PLN.
MA531 - Álgebra vetorial e linear para computação	Um dos modelos mais populares de PLN, word2vec, que trabalha com palavras em espaços vetoriais. Por isso, é preciso instrução relacionado a vetores e a espaços vetoriais.

### Referências

- [1] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit*. "O'Reilly Media, Inc.", 2009.
- [2] James H Martin and Daniel Jurafsky. *Speech and language processing*. Pearson/Prentice Hall, 2009.
- [3] Hinrich Schütze, Christopher D Manning, and Prabhakar Raghavan. *Introduction to information retrieval*, volume 39. Cambridge University Press, 2008.

<sup>1</sup>COLE, Arthur. *Natural Language Processing Turns to the Enterprise*. Disponível em: <<https://www.itbusinessedge.com/blogs/infrastructure/natural-language-processing-turns-to-the-enterprise.html>>. Acesso em: 07 de maio de 2018