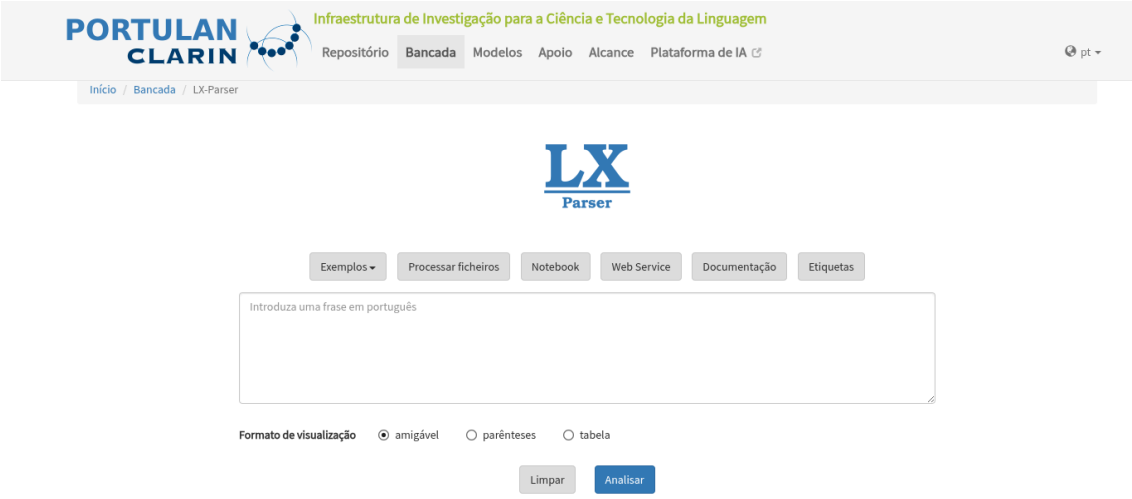
# Relatorio - Parsers

## Tipos de Parsers ou modelos (open souce) para substituição do PALAVRAS

### ▼ LX Parser (Universidade de Lisboa)




O LX-Parser foi desenvolvido com base no Stanford Parser. O analisador sintático desenvolvido pela Universidade de Stanford é um analisador sintático estatístico onde o treino é realizado com um *corpus* previamente anotado.

- **Sistema operacional**
  1. macOS / OS X
  2. Windows
  3. Linux

  Windows requer precisa de um software de compressão e descompressão, como os que são usados para arquivos pois o arquivo do parser esta disponibilizado em .gz
- **Linguagem de programação**
  - Python e Java

▼ **Documentação**

PORTULAN CLARIN / Workbench / LX-Parser

PORTULAN CLARIN: Research Infrastructure for the Science and Technology of Language

https://portulanclarin.net/workbench/lx-parser

**Autores**

LX-Parser – PORTULAN CLARIN

LX-Parser is a freely available on-line service for constituency parsing of Portuguese sentences. This service was developed and is maintained at the University of Lisbon by the NLX-Natural Language and ...

https://portulanclarin.net/repository/browse/lx-parser/893f597ed01511eb8cef02420a8701e72a135996a6bf498ba3813887aa3817ab

- **Informações sobre os modelos usados**
  1. **Albertina PT-BR (sem brWaC)**

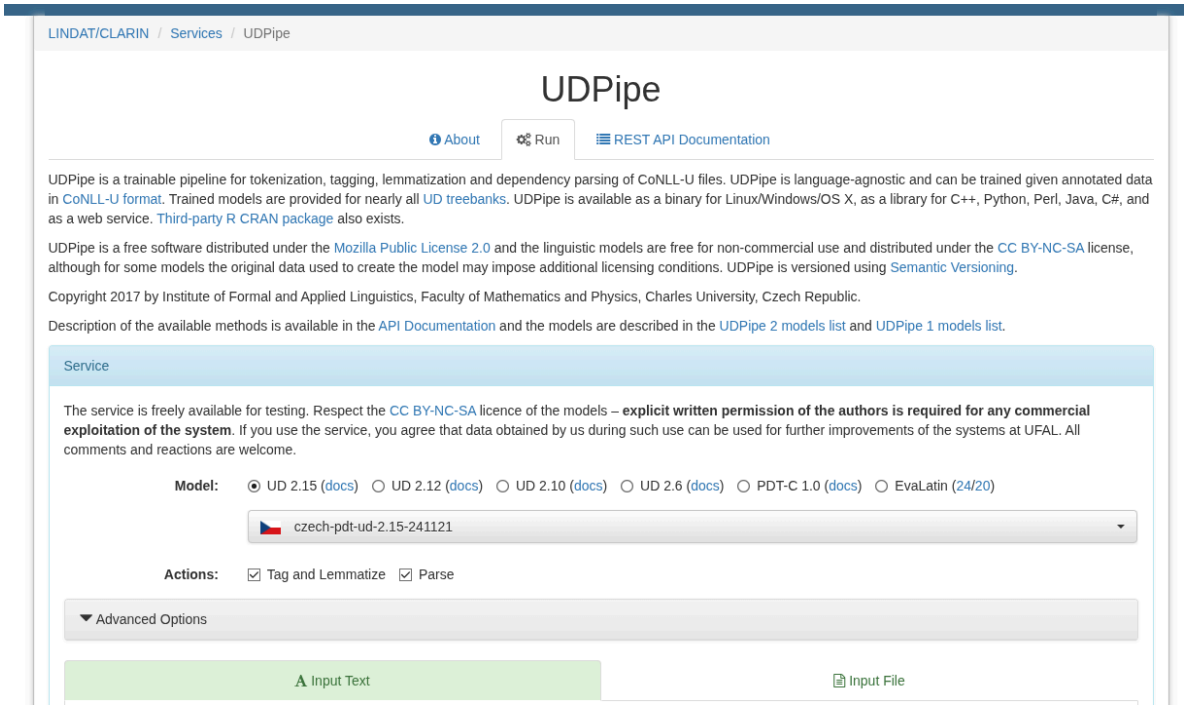
     Albertina PT-BR (sem brWaC) é uma versão para português americano do Brasil treinado em conjuntos de dados diferentes do brWaC e, portanto, com uma licença mais permissiva.

     Este modelo está disponível em três tamanhos, especificamente 1,5 mil milhões, 900 e 100 milhões de parâmetros. Visite as seguintes páginas do HuggingFace para obter instruções sobre como utilizar cada um destes modelos nas suas experiências:
  2. **Albertina PT-BR (brWAC)**

     Albertina PT-BR é a versão para português americano do Brasil, treinada no conjunto de dados brWaC.
  3. **BERTimbau Base**
  4. **BERTimbau Grande**

## Colab LX parser

## ▼ UDPipe (Universal Dependencies)



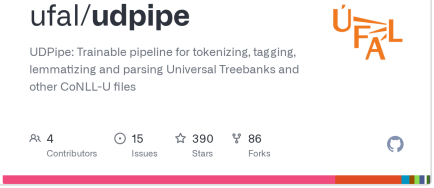

Grande conjunto que contém 147 modelos em 78 idiomas( No caso do português, o padrão utilizado é PT-PR). Cada um consistindo em um tokenizador, tagger, lematizador e parser, todos treinados usando os dados UD.

- **Sistema operacional**
  1. OS X
  2. Windows
  3. Linux
- **Linguagem de programação**
  - Python
- **Documentação**

GitHub - ufal/udpipe at udpipe-2

UDPipe: Trainable pipeline for tokenizing, tagging, lemmatizing and parsing Universal Treebanks and other CoNLL-U files - GitHub - ufal/udpipe at udpipe-2

https://github.com/ufal/udpipe/tree/udpipe-2



UDPipe

UDPipe is a trainable pipeline for tokenization, tagging, lemmatization and dependency parsing of CoNLL-U files. UDPipe is language-agnostic and can be trained given annotated data in CoNLL-U format. Trained models are provided for

https://lindat.mff.cuni.cz/services/udpipe/run.php

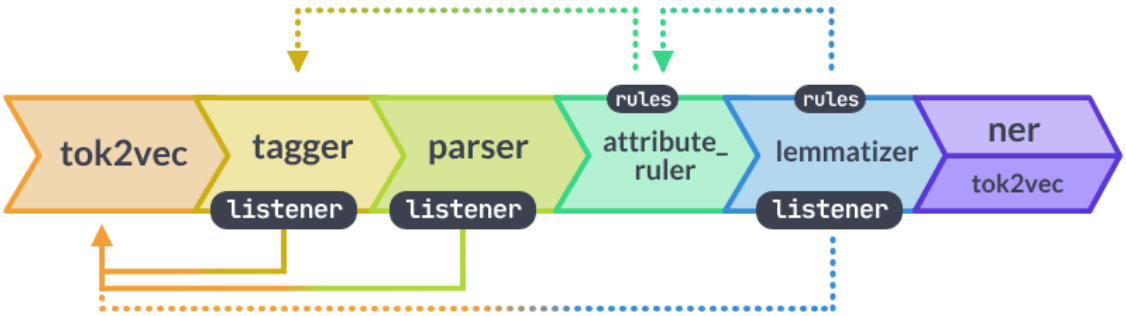
## Teste WEB

### ▼ spaCy

https://github.com/UniversalDependencies/UD_Portuguese-Bosque

Modelo base inspirado no PALAVRAS de pipelines treinadas em portugues, disponivel no spacy.

**Trained pipeline design**

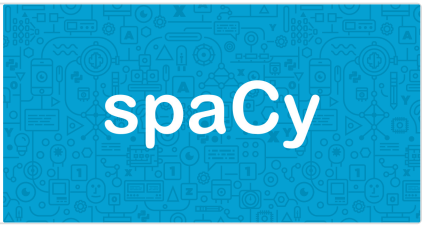



- **Sistema operacional**
  1. macOS / OS X
  2. Windows
  3. Linux
- **Linguagem de programação**
  - Python
- **▼ Documentação**

Linguistic Features · spaCy Usage Documentation

spaCy is a free open-source library for Natural Language Processing in Python. It features NER, POS tagging, dependency parsing, word vectors and more.

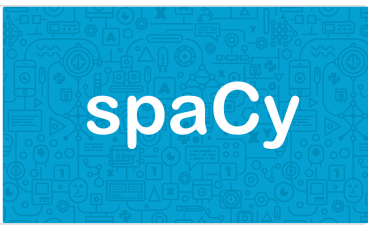
https://spacy.io/usage/linguistic-features#dependency-parse



spaCy · Industrial-strength Natural Language Processing in Python

spaCy is a free open-source library for Natural Language Processing in Python. It features NER, POS tagging, dependency parsing, word vectors and more.

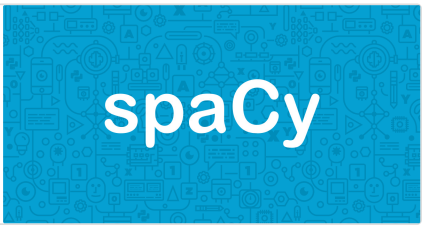
https://spacy.io/




Trained Models & Pipelines · spaCy Models Documentation

Downloadable trained pipelines and weights for spaCy

https://spacy.io/models

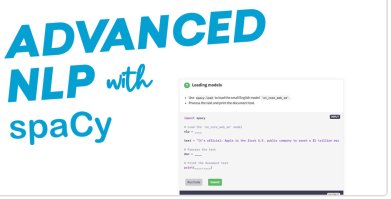



## Curso de spacy avançado

Advanced NLP with spaCy · A free online course

spaCy is a modern Python library for industrial-strength Natural Language Processing. In this free and interactive online course, you'll learn how to use spaCy to build advanced natural language understanding systems,

https://course.spacy.io/en




## Colab

## ▼ Stanza (Stanford NLP)

- **Sistema operacional**
  1. macOS / OS X
  2. Windows

  1. Linux
- **Linguagem de programação**
  - Python
- **Documentação**

Tutorials

High-performance human language analysis tools, now with native deep learning modules in Python, available in many human languages.

https://stanfordnlp.github.io/stanza/tutorials.html

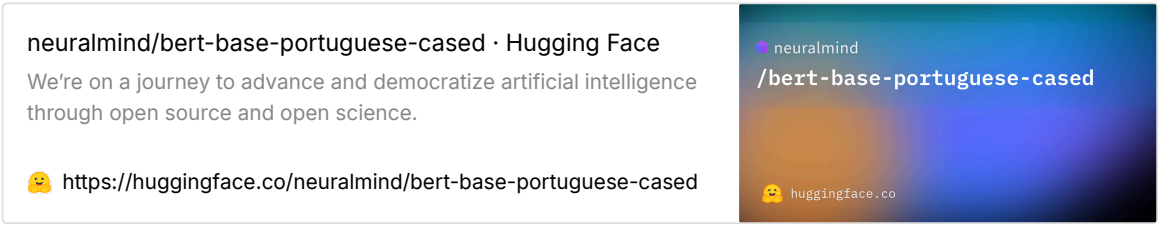
## Online Demo

### ▼ BERTimbau

O bert não tem o comportamento especifico de Parser, se trata de um modelo de LLM treinado em português com objetivos de ser uma ferramenta de apoio para aplicações de PLN

- **Sistema operacional**
  1. macOS / OS X
  2. Windows
  3. Linux
- **Linguagem de programação**
  - R e Python
- **Documentação**

  https://github.com/neuralmind-ai/portuguese-bert




## Colab

### ▼ PALAVRAS Parser

- **Sistema operacional**
  1. macOS / OS X
  2. Windows
  3. Linux
- **Linguagem de programação**
  - Java, Python, CMD line
- **Documentação**

VISL - PALAVRAS tag set

For more explanations and examples, as well as a comparison with the VISL form and function tags used in Palavras' graphical tree analysis, cf. Portuguese VISL category set.

https://edu.visl.dk/visl/pt/info/portsymbol.html

## 📊 COMPARAÇÃO RÁPIDA

| Parser | Facilidade | Velocidade | Precisão | Foco Linguístico | Português Específico |
|---|---|---|---|---|---|
| spaCy | ⭐⭐⭐⭐⭐ | ⭐⭐⭐⭐⭐ | ⭐⭐⭐⭐ | ⭐⭐⭐ | ⭐⭐ |
| Stanza | ⭐⭐⭐⭐ | ⭐⭐⭐ | ⭐⭐⭐⭐⭐ | ⭐⭐⭐⭐ | ⭐⭐⭐ |
| Palavras | ⭐⭐ | ⭐⭐ | ⭐⭐⭐⭐ | ⭐⭐⭐⭐⭐ | ⭐⭐⭐⭐⭐ |
| UDPipe | ⭐⭐ | ⭐⭐⭐ | ⭐⭐⭐⭐ | ⭐⭐⭐⭐ | ⭐⭐⭐⭐ |
| BERTimbau | ⭐⭐ | ⭐⭐ | ⭐⭐⭐⭐⭐ | ⭐⭐⭐ | ⭐⭐⭐⭐⭐ |
| LX-Parser | ⭐⭐ | ⭐⭐ | ⭐⭐⭐⭐ | ⭐⭐⭐⭐⭐ | ⭐⭐⭐⭐⭐ |

## 🔍 COMPARAÇÃO ESPECÍFICA

| Feature | PALAVRAS | Stanza | UDPipe | BERTimbau* | LX-Parser |
|---|---|---|---|---|---|
| Análise Morfológica | ✅✅✅✅✅ | ✅✅✅✅✅ | ✅✅✅✅⚪ | ✅✅✅⚪⚪ | ✅✅✅✅✅ |
| Lemmatização | ✅✅✅✅✅ | ✅✅✅✅✅ | ✅✅✅✅✅ | ✅✅✅⚪⚪ | ✅✅✅✅⚪ |
| Dependências | ✅✅✅✅⚪ | ✅✅✅✅✅ | ✅✅✅✅✅ | ✅✅✅✅✅ | ✅✅✅✅⚪ |
| Tags Linguísticas Finas | ✅✅✅✅✅ | ✅✅✅✅⚪ | ✅✅✅✅⚪ | ✅✅⚪⚪⚪ | ✅✅✅✅✅ |
| Foco em Linguística | ✅✅✅✅✅ | ✅✅✅✅⚪ | ✅✅✅✅⚪ | ✅✅⚪⚪⚪ | ✅✅✅✅✅ |
| Português Nativo | ✅✅✅✅✅ | ✅✅✅⚪⚪ | ✅✅✅⚪⚪ | ✅✅✅✅✅ | ✅✅✅✅✅ |
| Universal Dependencies | ❌ | ✅✅✅✅✅ | ✅✅✅✅✅ | ✅✅✅✅✅ | ✅✅⚪⚪⚪ |
| Facilidade de Uso | ✅✅⚪⚪⚪ | ✅✅✅✅⚪ | ✅✅⚪⚪⚪ | ✅⚪⚪⚪⚪ | ✅⚪⚪⚪⚪ |