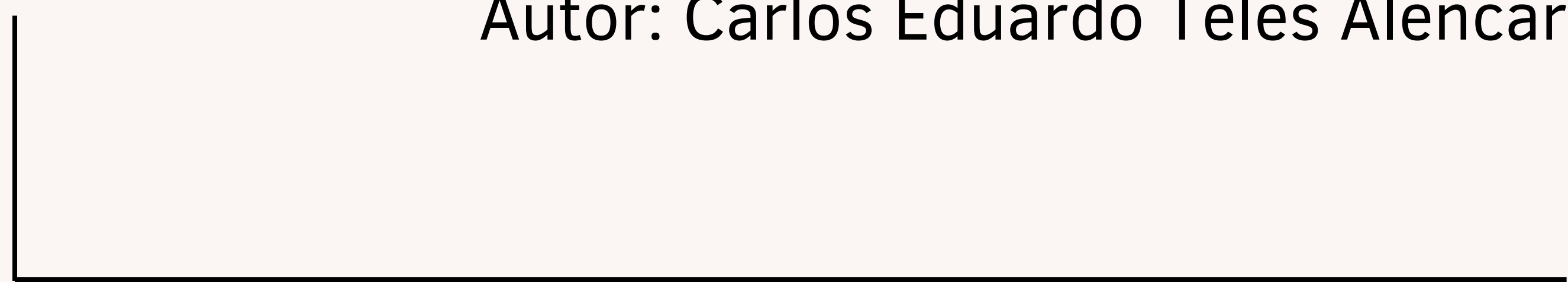




PROCESSAMENTO DE LINGUAGEM NATURAL COM STANZA

Autor: Carlos Eduardo Teles Alencar

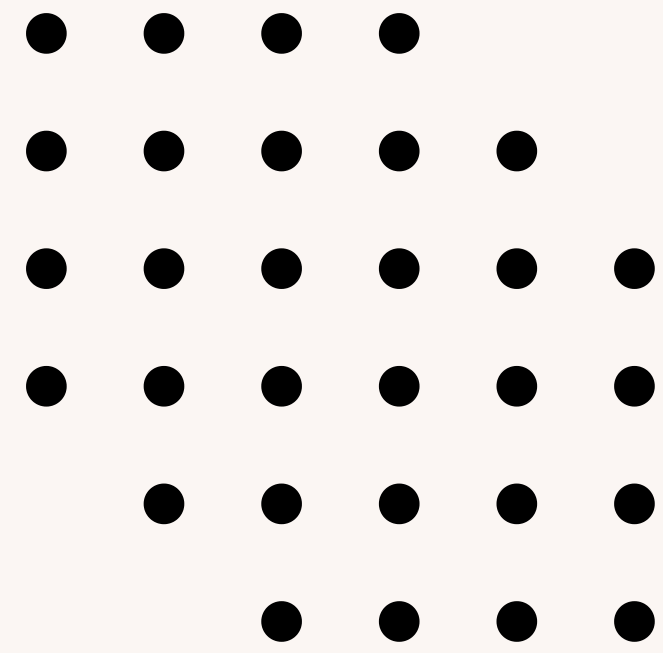


SUMÁRIO

- O que é Stanza
- Principais Tarefas
- Instalação
- Exemplos de uso
- Conclusão

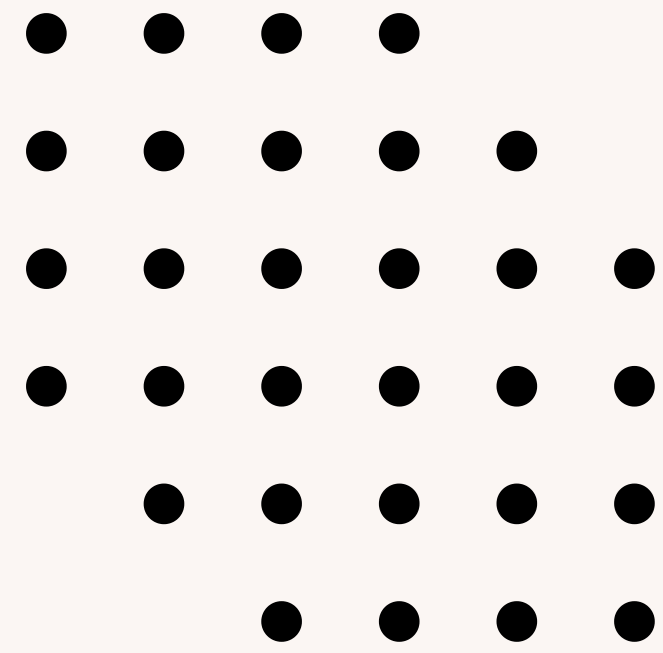
O QUE É STANZA

- Framework de análise linguística completa.
- Biblioteca desenvolvida em PyTorch
- Desenvolvida para ser precisa e multiplataforma.
- Mantida pela equipe da Stanford NLP Group.



PRINCIPAIS TAREFAS

- Tokenização
- Lematização
- POS-tagging
- Reconhecimento de Entidades
Nomeadas (NER)
- Análise de dependências
sintáticas



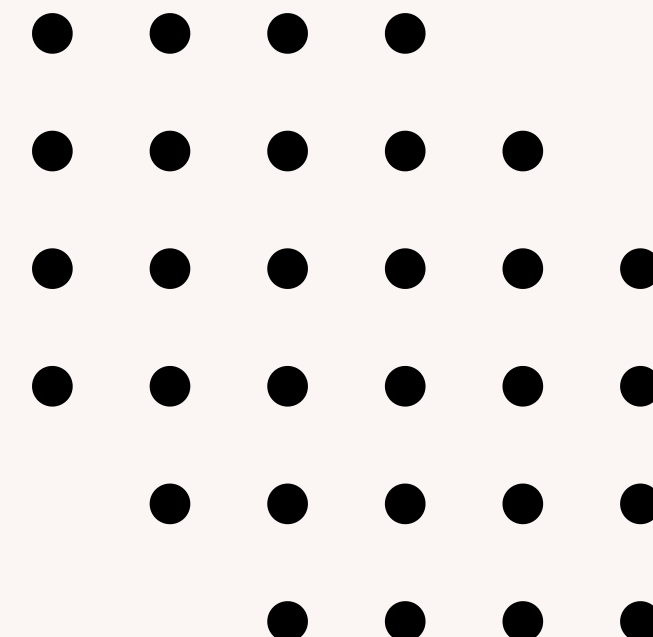
INSTALAÇÃO



```
pip install stanza
```



```
import stanza  
stanza.download('pt')
```



EXEMPLO DE USO

```
import stanza

# Carrega o pipeline de português
nlp = stanza.Pipeline(lang='pt', processors='tokenize,pos,lemma,ner')

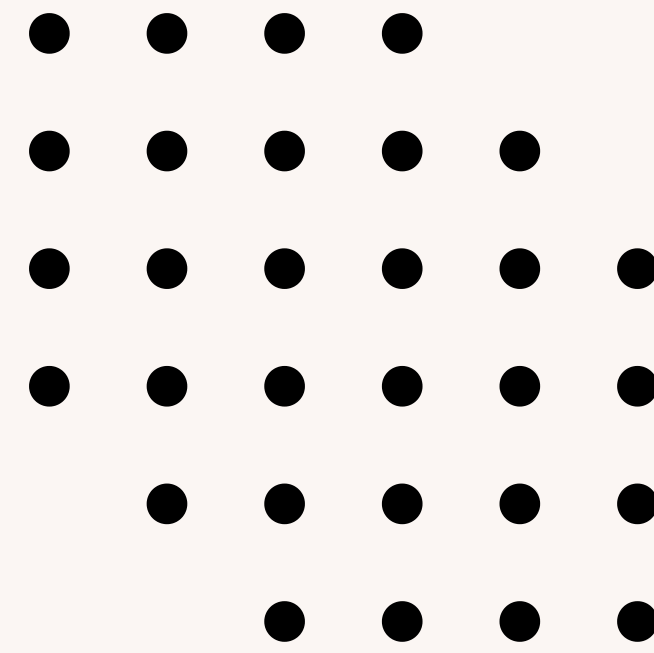
texto = "A Maria visitou Brasília em janeiro de 2025."
doc = nlp(texto)

for sent in doc.sentences:
    for word in sent.words:
        print(f"{word.text:<10} | {word.lemma:<10} | {word.upos}")
```

A		o		DET
Maria		Maria		PROPN
visitou		visitar		VERB
Brasília		Brasília		PROPN
em		em		ADP
janeiro		janeiro		NOUN
de		de		ADP
2025		2025		NUM
.		.		PUNCT

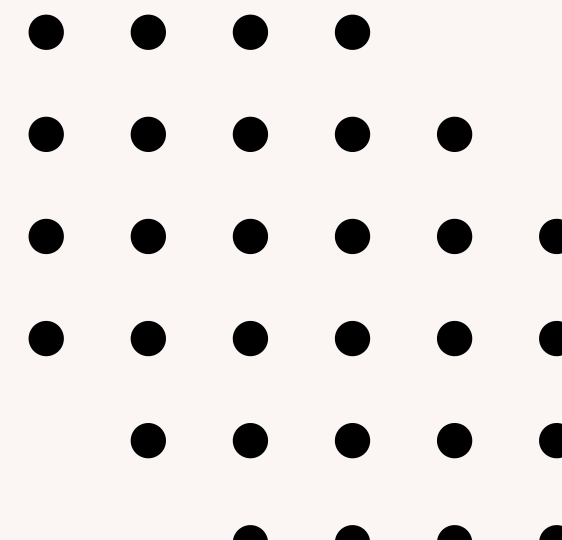
O QUE STANZA FAZ BEM

- Acurácia alta em
POS/lemma/dependências.
- Cobertura de idiomas ampla
com UD
- Morfologia detalhada
- Pipeline coerente “fim-a-fim”
pronto para uso



LIMITAÇÕES

- Mais lenta e pesada que o Spacy em geral
- Download de modelos por idioma (dezenas e centenas de MB)
- Treino/customização menos “ergonômicos” que o Spacy
- Não corrige gírias/gramática



COMPARATIVO RÁPIDO

Aspecto	Stanza	Spacy	Enelvo
Foco	Acurácia e UD	Velocidade e produção	Normalização PT informal
Tarefas	Tokenize, POS, lemma, dep, NER	tokenize, POS, lemma, dep, NER , train, CLI, ecosystem	Corrige gírias/abreviações/erros comuns
Velocidade	Mais lento	Mais rápida	n/a (pré-processamento)
Idiomas	Muitos (UD)	Muitos (modelos próprios)	Português
Customização	Média (Scripts)	Alta (config/training, extensões)	Dicionário/heurísticas
Correções gramaticas	Não	Não	Não

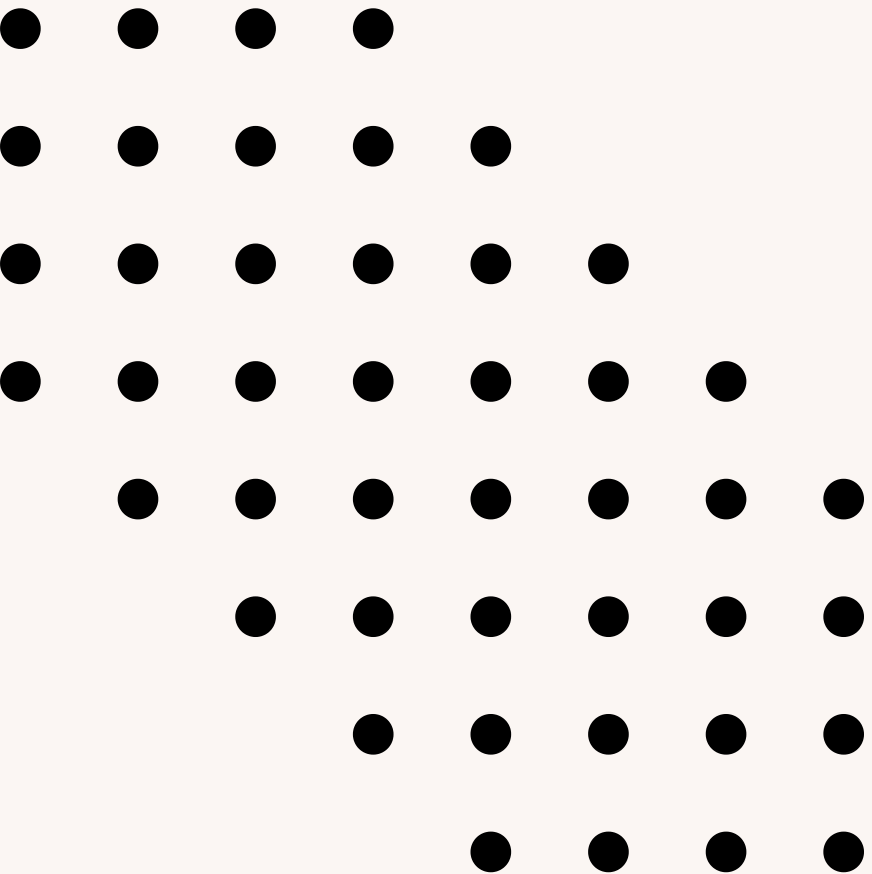
CONCLUSÃO

- A Stanza é uma alternativa robusta ao spaCy, com forte suporte linguístico e modelos treinados em português.
- Pode ser combinada com Enelvo (para limpeza de texto) e spaCy (para aplicações mais rápidas).
- Usado para análise morfossintática detalhada



TABELA DAS TAGGINGS E SUAS EXPLICAÇÕES

Tag	Classe Gramatical	Explicação	Exemplos
ADJ	Adjetivo	Palavra que caracteriza um substantivo, indicando qualidade, estado ou condição.	bonito, alto, inteligente
ADP	Adposição (Preposição)	Palavra que liga termos, indicando relação de tempo, lugar, causa, modo, etc.	de, em, com, para
ADV	Advérbio	Modifica o sentido de verbos, adjetivos ou outros advérbios.	rapidamente, bem, ontem
AUX	Verbo auxiliar	Acompanha o verbo principal para formar tempos compostos, vozes ou modos verbais.	está, foi, tinha
CCONJ	Conjunção coordenativa	Conecta orações ou termos de mesma função.	e, mas, ou
DET	Determinante	Acompanha o substantivo, indicando gênero, número ou posse.	o, a, os, as, este, uma
INTJ	Interjeição	Expressa emoções, sentimentos ou reações.	ah, nossa, olá, opa
NOUN	Substantivo comum	Nomeia seres, objetos, ideias, sentimentos, lugares, etc.	casa, carro, pessoa
NUM	Numeral	Expressa quantidade, ordem, multiplicação, fração, etc.	um, dois, primeiro
PART	Partícula	Termos que não se encaixam nas demais classes, geralmente formam locuções ou expressões fixas.	se, que (em certas construções)
PRON	Pronome	Substitui ou acompanha o substantivo, indicando pessoas, coisas ou posse.	eu, você, ele, nós
PROPN	Substantivo próprio	Nomeia seres ou entidades únicas, sempre com inicial maiúscula.	Maria, Brasil, Google
PUNCT	Pontuação	Representa sinais gráficos de pontuação.	., ?, !, ,
SCONJ	Conjunção subordinativa	Introduz orações subordinadas, estabelecendo relação de dependência.	porque, embora, se
SYM	Símbolo	Representa símbolos matemáticos, monetários ou especiais.	\$, %, +, =
VERB	Verbo principal	Indica ação, estado ou fenômeno.	correr, pensar, existir
X	Outro / Desconhecido	Palavras que o modelo não reconhece ou não classifica (estrangeirismos, gírias novas, erros).	OK, lol, hashtag



Obrigado

