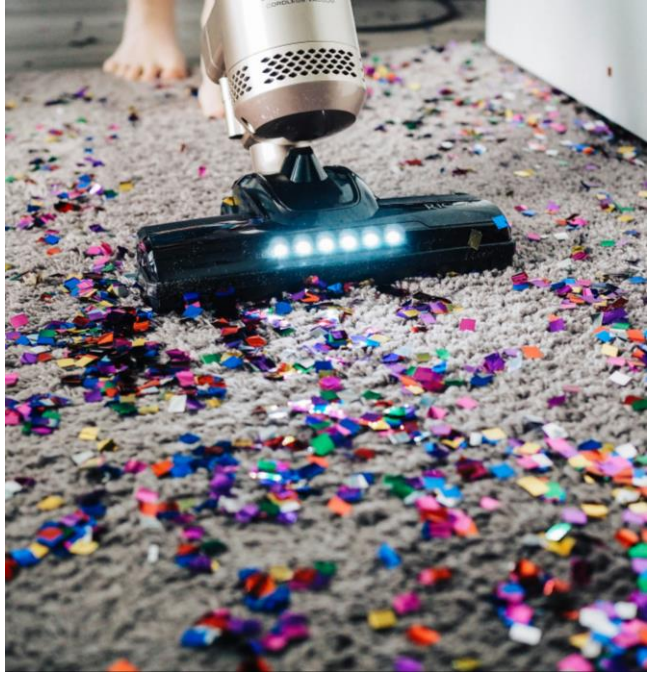
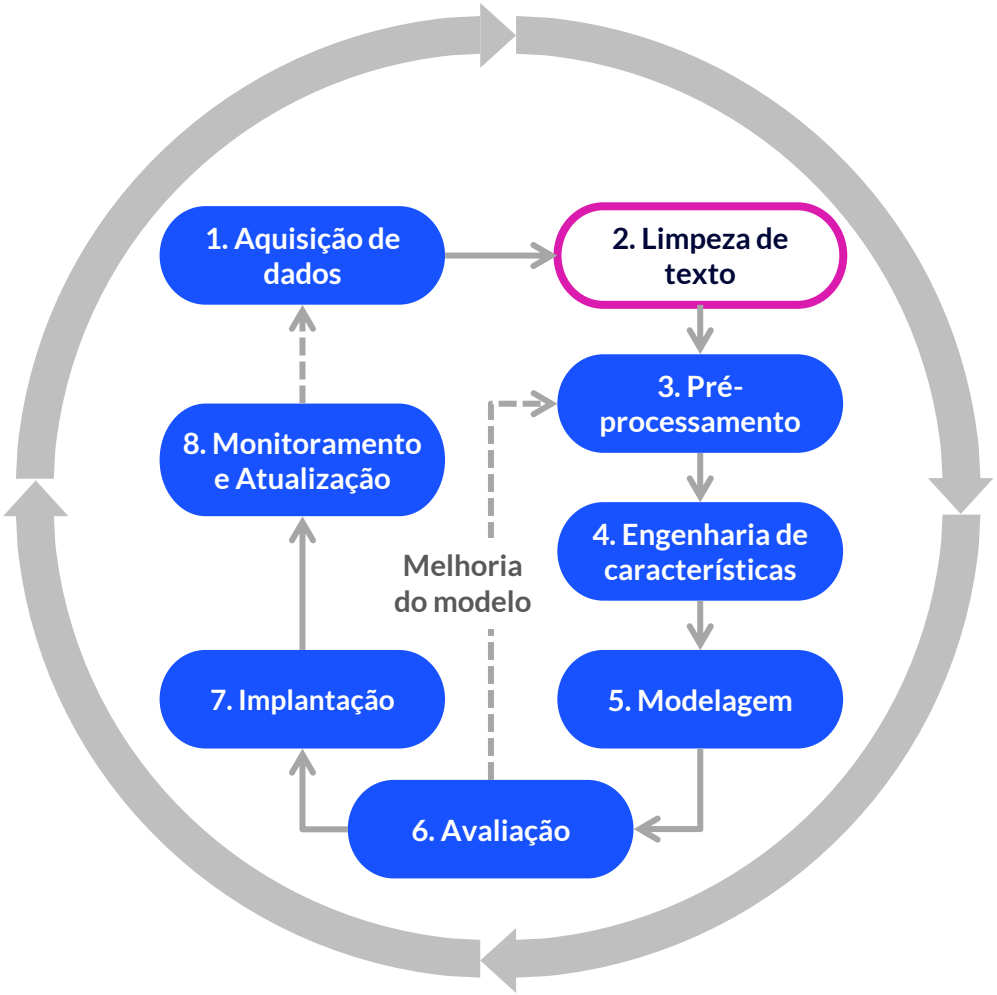
PUC Minas Virtual

PROCESSAMENTO DE LINGUAGEM NATURAL

LIMPEZA DE DADOS

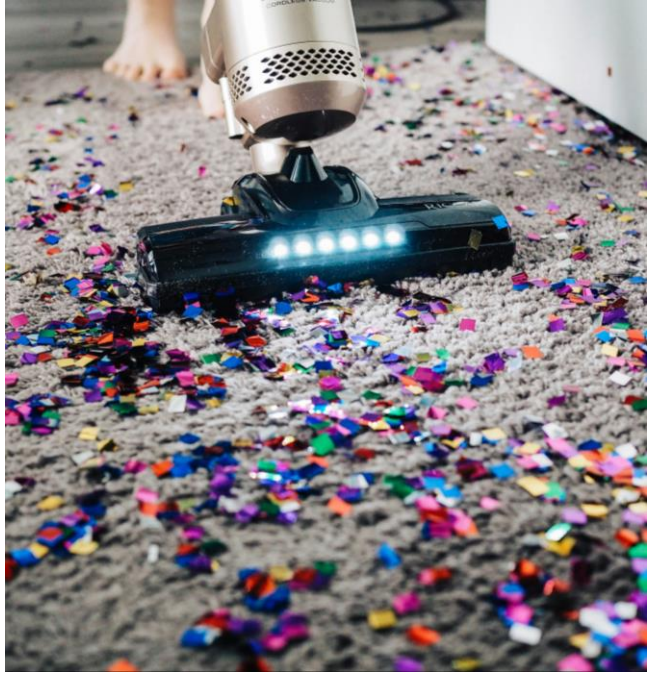


Limpeza de Dados



Limpeza de Dados

Extração do texto bruto dos dados de entrada, removendo informações não textuais



Limpeza de Dados

01 Remoção de dados não textuais

02 Correção ortográfica

03 Normalização Unicode

Remoção de dados não textuais

Tipos de Documentos

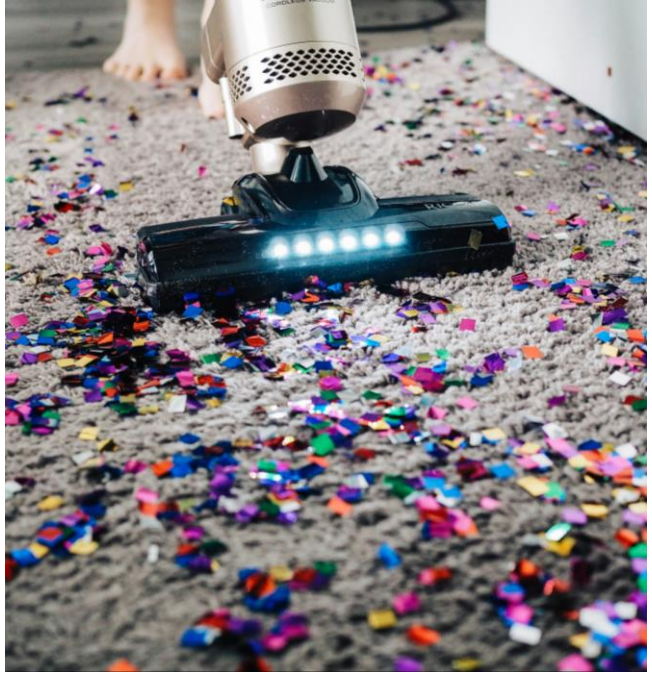
Portable Document Format

HyperText Markup Language

Texto Embutidos em Imagens

BeautifulSoup

https://pypi.org/project/beautifulsoup4/



Limpeza de Dados

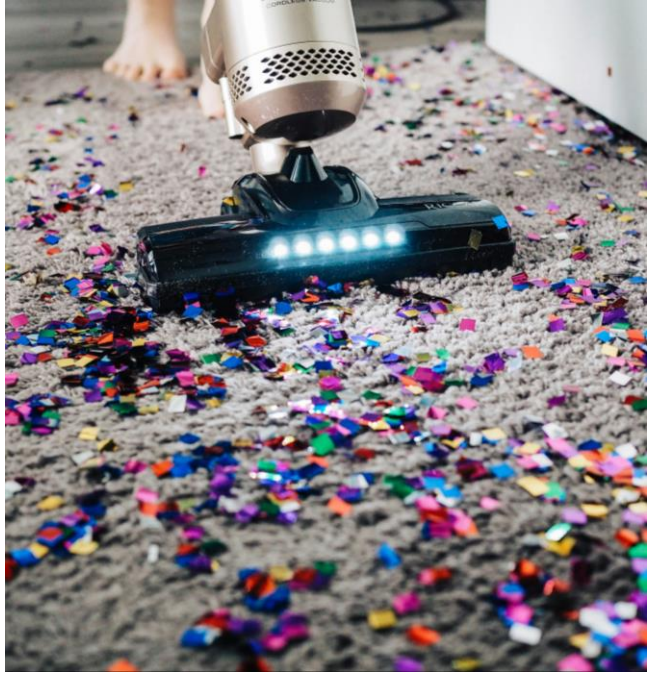
01 Remoção de dados não textuais

02 Correção ortográfica

03 Normalização Unicode

Spellchecker

https://pypi.org/project/pyspellchecker/



Limpeza de Dados

01 Remoção de dados não textuais

02 Correção ortográfica

03 Normalização Unicode

encode()

PUC Minas
Virtual