

03

Spacy: Doc e Tokens

Nesta aula, exploramos melhor a biblioteca **Spacy** e a utilizamos para nos ajudar no pré-processamento dos dados textuais.

A função que faz o pré-processamento removendo stop words, caracteres não alfabéticos e retornando apenas títulos com mais de 2 palavras é a `trata_textos`.

Dentre as opções a seguir, qual implementa a função `trata_texto` corretamente?

Selecione uma alternativa

A

```
def trata_textos(doc):
    tokens_validos = []
    for token in doc:
        e_valido = not token.is_stop and token.is_alpha
        if e_valido:
            tokens_validos.append(token.text)

    if len(tokens_validos) > 2:
        return " ".join(tokens_validos)
```

B

```
def trata_textos(doc):
    tokens_validos = []
    for token in doc:
        e_valido = token.is_stop and token.is_alpha
        if e_valido:
            tokens_validos.append(token.text)

    if len(tokens_validos) > 2:
        return " ".join(tokens_validos)
```

C

```
def trata_textos(doc):
    tokens_validos = []
    for token in doc:
        e_valido = not token.is_stop and token.is_alpha
        if e_valido:
            tokens_validos.append(token)

    if len(tokens_validos) > 2:
        return " ".join(tokens_validos)
```