



O processo de stemming

O conceito de stemm

Stemm é uma palavra do inglês que significa caule e uma parte de uma planta que liga as folhas da planta.

No âmbito do processamento de linguagem natural, stemm identifica o radical das palavras, que é aquela parte mantida independente das variações de gênero ou número.

Análise léxica

Para a realização das análises léxicas é importante que se reconheça o lemma e suas variações.

A extração do radical pode ser crucial nessas situações, porque extraindo o radical de duas palavras é possível compará-las e identificar se elas possuem a mesma origem.

O processo de stemming

O processo de stemming consiste em reduzir a palavra em questão para o seu radical, reduzindo a complexidade das palavras ao transformando em seu lema canônico.

Esse processo é executado durante a etapa de análise léxica, facilitando a identificação dos sentidos das frases e também reduzindo o tamanho dos dados, facilitando a realização de um trabalho mais rápido.

O processo de stemming

O processo de stemming



Se considerássemos a palavra CONTROLE para o processo de stemming, ao final seria obtido CONTROL.

Essa operação é executada em todas as palavras.

O processo de stemming

A biblioteca *nltk* possui um algoritmo implementado para a realização do processo de stemming. Para utilizá-lo é necessário importar a biblioteca e essa implementação.

```
import nltk  
from nltk.stem import RSLPStemmer
```

O processo de stemming

O processo de stemming

A utilização do algoritmo de stemming é bastante simples:

```
stemer.stem("Entregar")
```



Obrigada!

Ana Laurentino