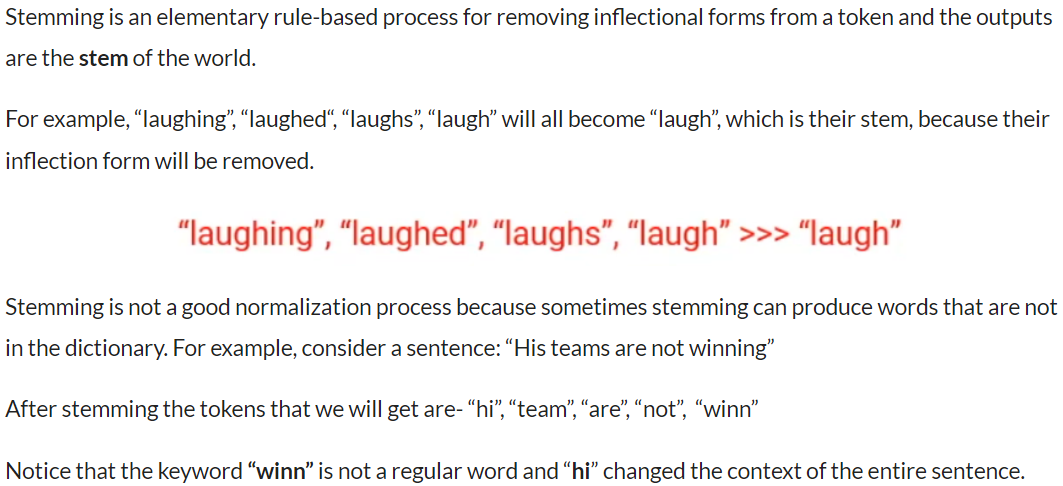
Tokenization:

-White-space Tokenization

-Regular Expression Tokenization

Normalization:

-Stemming



-Lemmatization

