

I saw an amazing thing $\xrightarrow{\text{stem}}$ I s an amazing thing

I saw an amazing thing $\xrightarrow{\text{lemma}}$ I see an amazing thing

- We need to **normalize** text in order to analyze it.
- This means we want all the words to be lowercased or to convert plural terms into singular ones. We often do this with two methods:
 - **Stemming** chops off the ends of words in the hope of achieving this.
 - **Lemmatization** instead returns the base or dictionary form of a word, which is known as the lemma. It's the more robust approach than stemming.

THE N-GRAM

