

Summary

- We can think of text as a sequence of tokens
- Tokenization is a process of extracting those tokens
- We can normalize tokens using stemming or lemmatization
- We can also normalize casing and acronyms
- In the next video we will transform extracted tokens into features for our model