#############################################################  
# File: Homework4\_Narrative\_elo180000.word #  
# Author: Ethan Ong (ELO180000) #  
# Date: 3/3/2023 #  
# Purpose: CS 4395 - Homework 4 (N-grams Language Models) #  
#############################################################

1. **What are n-grams and how are they used to build a language model**

N-grams is a sliding window of size n over text. N-grams create a probabilistic model of language.

1. **List a few applications where n-grams could be used**

Some applications where n-grams could be used are speech recognitions and text summarization.

1. **A description of how probabilities are calculated for unigrams and bigrams**

To calculate the probability of unigrams you would use the number of occurrences of a token in a text and divide it by the total number of tokens. For bigrams, you take the probability of the first word and multiply it by the probability of the second word.

1. **The importance of the source text in building a language model**

When you have a large source text, the language model can become better. If you have a small source text the language model would not be as good. Plus the bigger it is the more accurate the model will be.

1. **The importance of smoothing, and describe a simple approach to smoothing**

Smoothing in n-grams are important since it helps with any outliers in the sets. It replaces zero counts/probabilities with counts/probabilities of words that occur only once. An example of smoothing is laplace smoothing. Laplace smoothing only adds 1 to the count of each word and bigram.

1. **Describe how language models can be used for text generation, and the limitations of this approach**

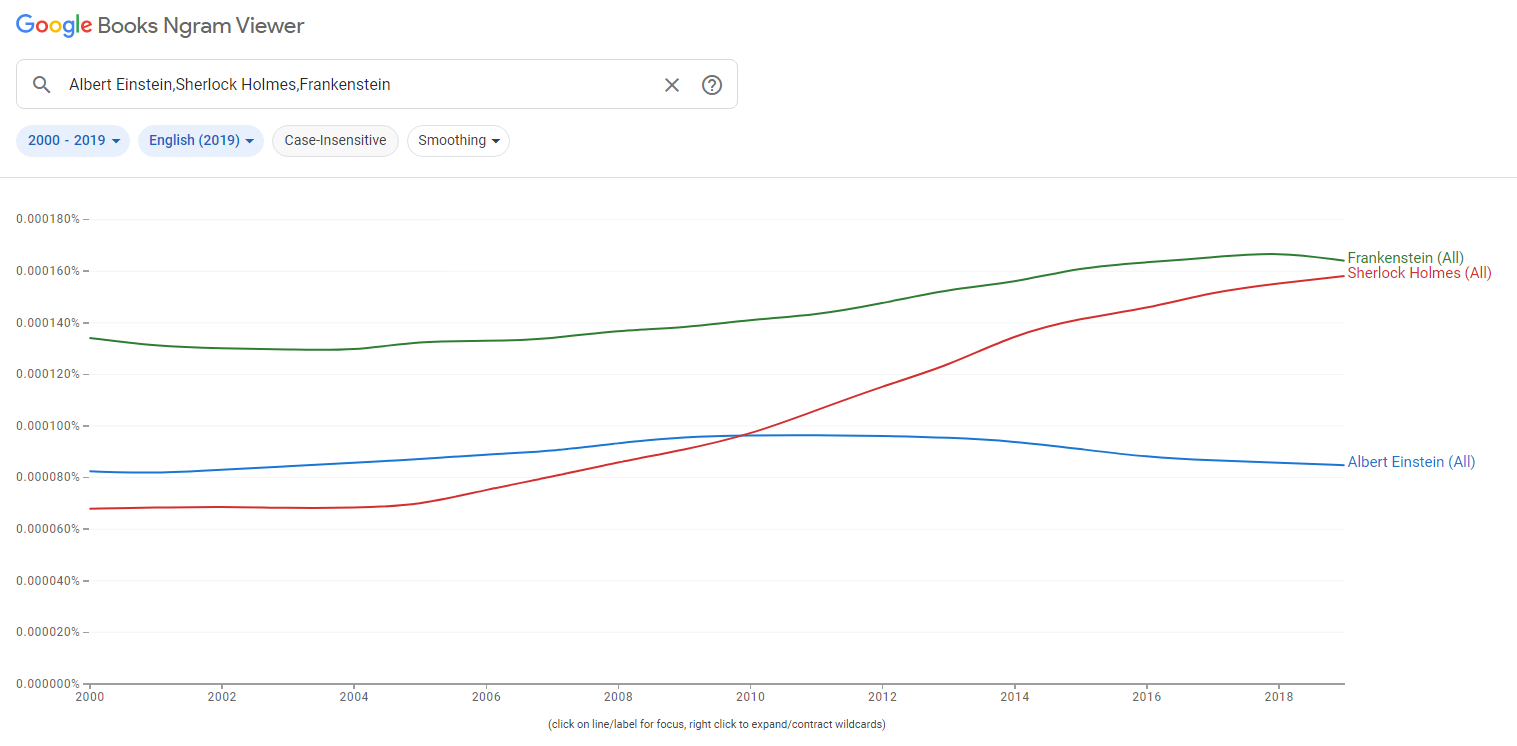
Language Models can be used for text generations by generating dictionaries of all the words and the probabilities of each words. The limitation depends on how big the source text is in the language model. If it is really large it can be really good, however, if the source text is small it would not be as good and it would be not a accurate.

1. **Describe how language models can be evaluated**

Language models can be evaluated by extrinsic: human annotators or intrinsic: an internal metric like perplexity

1. **Give a quick introduction to Google’s n-gram viewer and show an example**

This is an example from Google’s n-grams on Albert Einstein, Sherlock Holmes, and Frankenstein

****