

* n-grams

find frequency of 'th' \Rightarrow 2 gram

Pass a 'n' gram, find total frequency

Plot cells on x and y axis

numpy documentation? To understand, visualize
 \hookrightarrow matrix operations.

* Features of text

- Attempt to analyze text into numbers.
- Also called as feature extraction / vectorization

Compare histograms, unigrams (Mostly similar)
 \Rightarrow If similar, then not a useful method

Example of 2 wikipedia pages on Giraffe and elephant, one translated in English other in French.

~~Specifically~~ Specifically how do we know optimal value of n in 'n grams'

* Written Numbers

- \hookrightarrow Extract features from images
- \hookrightarrow Each image is 28×28 array

Scatter plot - 3D

Technique: Count number of black pixels that are surrounded on four sides by non-black pixels.