**Document wide statistics:**

- Total number of tokens

- Top 3 most common tags

- Top 3 most common tokens

- Top 3 most common lemmas

**Statistics for search of token:**

1. number of appearances in text out of total number of words in text:

publicString FreqOfToken(String aWord)

2. Part of Speech Tags and Percentage of the POS in the text:

**public** String FreqOfCertainTags(String aWord)

3. Most likely to be the next word Top 5 (hello 6% of the time there 7% of the time)

**public** String getprevWord(String aWord)

To do: display only top 5 words sorted by number of appearance

What if word is at end of sentence? Beginning?

4. Most likely to be the next POS Tag Top 3 (NN 35% of the time VB 60% of the time...)

5. Most likely to be the previous word Top 3 (hello 5% of the time there 8% of the time)

**public** String getfollowingWord(String aWord)

To do: display only top 5 words sorted by number of appearance

What if word is at end of sentence? Beginning?

6. Most likely to be the previous POS Tag Top 3 (NN 35% od the time PP 7% of the time)

optinonal/ideas:

- beginning, middle or end of the document + most common

- beginning, middle or end of the sentence + most common

- number of letters + is it a long or a short word + frequency of long words + short words

**Statistics for search of POS TAG:**

- Frequency of that tag inside the document

- Top 5 most common words with that Tag:

- Top 5 most common lemmas with that Tag:

- Most likely to be the next POS (Top3):

- Most likely to be the previous POS (Top 3):

**Statistics for search of lemma:**

- Frequency of that lemma inside the text

- tokens that have that lemma + count

- different POS of that lemma

- which POS is the most common with that lemma