I. Basic Tools for Textual Analysis (+ Part-Of-Speech & Named Entity Recognition)
   a. **Voyant Tools**: [voyant-tools.org](http://voyant-tools.org) (word frequencies, word clouds, KWIC)
      
      *Python commands* (NLTK: Text object; collocations, KWIC, word frequencies): 
      Basic Text analyses.ipynb
   
   b. **POS & NER**: stanford-postagger-3.7.0.jar, stanford-ner-3.7.0.jar
      
      List of POS tags: 
      [https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html](https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html)
      
      *Python (SpaCy) POS & NER*: POS-tagging and Lemmatization in SpaCy.ipynb
      NER in SpaCy.ipynb
      
      SpaCy installation instructions: [https://spacy.io/usage](https://spacy.io/usage)
      SpaCy NER tags : [https://spacy.io/usage/linguistic-features](https://spacy.io/usage/linguistic-features)
      
      HathiTrust Research Center: [analytics.hathitrust.org](http://analytics.hathitrust.org)
      HTRC Bookworm (Ngram search): [https://bookworm.htrc.illinois.edu/develop/](https://bookworm.htrc.illinois.edu/develop/)
      
      "Visual Text Explorer": [edoc.uchicago.edu/vte](http://edoc.uchicago.edu/vte)  “simultaneous close and distant reading”

II. Tools for Stylometry (HCA Dendogram & k-means PCA)
   a. **LEXOS** (Comparative Stylometry: Dendrogram + PCA): [lexos.wheatoncollege.edu](http://lexos.wheatoncollege.edu)
   
   b. **Python-based Stylometry**: Stylometry_HCA.ipynb, Stylometry_PCA.ipynb

III. Tools for Topic Modeling + Word2vec
   a. **MALLET Topic Modeling**: [mallet.cs.umass.edu](http://mallet.cs.umass.edu)
      TopicModelingTool.jar: standalone Java-based application for Topic Modeling
      
      b. **Python-based Topic Modeling** (via the gensim library, NLTK + SpaCy): 
         Topic Modeling (gensim LDA + NLTK + SpaCy)_Shakespeare.ipynb
         Topic Modeling evaluations_Shakespeare.ipynb
      
      c. **Python-based Word2vec & TF-IDF** (gensim): Word2Vec all_Shakespeare.ipynb
         Word2Vec TF-IDF Shakespeare.ipynb

IV. Tools for Text Reuse
   a. **Philologic4**: [http://anomander.uchicago.edu/text-pair/](http://anomander.uchicago.edu/text-pair/)
      [https://textual-optics-lab.uchicago.edu/](https://textual-optics-lab.uchicago.edu/)
      
      PCA + BLAST (for genomic/literary sequence analysis): [pvierth/herokuapp.com](http://pvierth/herokuapp.com)
Instructions for counting the POS tags from the stanford-postagger using Word + Excel:

1. Open the stanford-postagger-3.7.0.jar
2. Paste in your text, click “Tag Sentence!”
3. Copy the output, go to Microsoft Word, paste the output
4. In Word:
   a. Go to “Advanced Find & Replace”
   b. Find:  [enter one space, it will be invisible]
   c. Replace: ^p [the paragraph marker]
   d. Select all, Copy the entire document
5. In Excel:
   a. Paste in the text from Word (all the data will be in one column)
   b. Under the Data menu, click on “Text to Columns”
   c. Choose “Delimited”, click Next
   d. In the “Other” box, enter the underscore: _
   e. Click Finish, each word will be in Column A and each tag will be in Column B
   f. To do counts of parts of speech, copy the column with the tags (e.g. B)
   g. Paste it into a different column (e.g. D)
   h. Highlight the column, and under the “Data” tab, select “Remove Duplicates”
   i. Click “Remove Duplicates”
   j. In cell E1, enter the following formula: =COUNTIF(B:B,D1)
   k. Copy that cell to all the rest of the E cells where D has a tag
   l. Sort by column E (descending) and then by column D (A to Z)