Required Programs:

1. Text Processing and Extraction
   * Converts article text to processable format for data training and categorization (ie: text document with delimiters)
   * Should take .docx file path as argument and output file path as 2nd argument
   * Should not apply stemming or lemmatization, but removes special characters and formats
2. Training Program
   * Accepts a training and validation data set
   * Applies stemming and/or lemmatization to improve algorithm performance
   * Saves training data to file so it is reusable
3. Categorization Program
   * Accepts training data and data set
   * Processes, categorizes, and spits out an excel compatible csv file
   * GUI would be preferable, but not required (Optional)

Python Modules:

* TextBlob: Text Processing library for applying stemming and lemmatization as well as text structure recognition
* Python-Docx: Module to open, save, and process .docx files from Microsoft Word
* Numpy & Scipy: Modules for math and data processing
* Doc2text: Module specializing in plaintext extraction from docx files from Microsoft Word