1. Get ‘documents’
2. Tokenize documents, remove common words, words that only appear once
   1. Self-determine common words, e.g ‘the’
   2. Take array of not common words and use a dictionary to determine their frequency (word: frequency as key: value)
   3. Save array of words that appear more than once (in all documents)
   4. Genism example splits by whitespace and lowercases everything as done in Deerwester et all’s article (simple but not the most efficient) - <https://radimrehurek.com/gensim/tut1.html#id3>