## Homework 6

1. Evaluate text similarity of Amazon book search results by doing the following:
   1. Do a book search on Amazon. Manually copy the full book title (including subtitle) of each of the top 24 books listed in the first two pages of search results.
   2. In Python, run one of the text-similarity measures covered in this course, e.g., cosine similarity. Compare each of the book titles, pairwise, to every other one.
   3. Which two titles are the most similar to each other? Which are the most dissimilar? Where do they rank, among the first 24 results?
2. Now evaluate using a major search engine.
   1. Enter one of the book titles from question 1a into Google, Bing, or Yahoo!. Copy the capsule of the first organic result and the 20th organic result. Take web results only (i.e., not video results), and skip sponsored results.
   2. Run the same text similarity calculation that you used for question 1b on each of these capsules in comparison to the original query (book title).
   3. Which one has the highest similarity measure?

Submit all of your inputs and outputs and your code for this assignment, along with a brief written explanation of your findings.