* Start with at least 5 documents, text files (Get them by crawling the web, or by extracting some twitter data.  If twitter, you will need more than 5 tweets, but can make a "document" that is a collection of tweets on a topic or from a particular person or with a particular hashtag.
* Extract all the words in all the documents.
* Decide which words should appear in your dictionary.  (How will you decide that?  Will you have a stop word list or not? Will you eliminate words that contain non-alphabetic characters?)  (Be sure to ignore upper/lower case distinctions.)
* Build an index.  Will you choose to keep track of how many times each term appears in each document?  What effect do you expect that to have on the value of your index?
* Build a minimal search engine:
  + Allow a user to enter a search query
  + Find the documents that best match that query
  + Print out the IDs of the matching documents.

With such a small sample of documents, you need to restrict your searches to queries that are relevant to those documents.  The point of this is to bring together the concepts we have built during the first half of the semester so that you have a concrete understanding of what is involved in creating an information retrieval system.