Suppose that you’re constructing a paper on Internet usage and want to examine the various types of abstracts and the language usage therein. You could assemble a mini-corpus for exactly that purpose with the following steps:

1. Collect relevant articles—just a dozen or so will get you started.



2. Copy and paste the abstract of each article into a text file (UTF-8 encoding works best). Notice that I cleverly labeled each text file as just the “abstract”.



3. Simply analyze these files using some handy software such as WordSmith or AntConc.

4. Easy, wasn’t that?