Any data file which lists each unique word and all the places where that word may be found in a file or corpus of files is known as a "concordance". There are many corpus-oriented perspectives for which a concordance-building tool is very useful. I have posted a slightly-modified version of a [Python concordance-generating script](https://mycourses.rit.edu/d2l/common/dialogs/quickLink/quickLink.d2l?ou=766919&type=coursefile&fileId=_con.py) for you to look at, and using *macbeth.txt* for input gives us [macbeth.txt.concordance.txt](https://mycourses.rit.edu/d2l/common/dialogs/quickLink/quickLink.d2l?ou=766919&type=coursefile&fileId=macbeth.txt.concordance.txt) as output.

Note also the changes I made to that script:

1. the *input* file can now have *any* name, and
2. the *output* lines no longer start with a sequence number

The output format is now: one line per unique word (or record). Each line starts with that unique word (key), followed by whitespace, then by a sequence of one or more line numbers upon which that unique word may be found, and each line (record) terminating with a line-ending.

In no more than 250 words, write a response to the following:

Propose a different concordance output format that would allow rapidly locating or counting the instances of a given word appearing in *any* of *many* documents, each of which has it's own unique name/location. Do not just propose the technical details (data fields, separator character, etc), but also explain why you made the choices you did.

Upload/submit your response as a single file with both the technical details and the explanation to the dropbox.

In this day and age where nearly all text is digitized (see Google Books), I would create a fully hyperlinked concordance. Each key word would link directly to a definition (dictionary.com for example) and each specific instance would directly link to the word location within the text. It may require additional work to hyperlink the source material, but one could programmatically do such a task and systematically create a fragment identifier for every word. Then, any user of the concordance could simply click on the specific instance of the word within the concordance and be directly linked to it within the text.

This could be done in JSON, since JSON allows for an arbitrary number of entries, as opposed to a flat file which requires you to know ahead of time how many columns you will need. One could also easily get a count of a word in a single text or all included texts by aggregating each individual text (see below). Since we are looking at gathering many documents, the JSON could look like this:

* word (hyperlinked to dictionary entry)
  + text 1 (e.g. book it is in)
    - word 1 (hyperlinked to exact location within text 1)
    - word 2 (hyperlinked…)
  + text 2 (e.g. article it is in)
  + etc.