

SE375 - Laboratory Assignment 08

Word Count and Inverted Index: The Networking Version

Consider the Word Count and Inverted Index applications. This time, the processing will be coordinated by a server. The server will start with n number of URLs and it will wait for n number of connections from client applications. Once all of the clients are connected, the server will send each of them a URL for the file they will work on.

The clients, after receiving the URL for the file to work on, will download the file and start sending the results to the server. Once all of the words are complete, they will quit.

You can (and should) use the following URLs for your server.

- <https://www.cs.umb.edu/~smimarog/textmining/datasets/r8-train-all-terms.txt>
- <https://www.cs.umb.edu/~smimarog/textmining/datasets/r8-test-all-terms.txt>
- <https://www.cs.umb.edu/~smimarog/textmining/datasets/r52-train-all-terms.txt>
- <https://www.cs.umb.edu/~smimarog/textmining/datasets/r52-test-all-terms.txt>

In the previous versions of these applications, the user was able to ask for a word and the application would return the word count once the processing was complete. Now, the server should be able to answer such queries while the clients are working on the files. Since these are tiny requests, you can use a UDP client to query words and the server would return their word count.

