Attached is an archive of simple text files (documents), each contains one poem. The task is to find similarities between the documents.

Please write some code that given a document will return a list of n most similar documents. You may define similarity however you choose to, and use whichever programming language you feel most comfortable with.

We expect you to spend about an hour on this task and definitely no longer than 3 hours, and so it is understood that the solution will be a simple first-attempt prototype.

Once you are done with coding, please explain briefly the approach you chose, and describe how you would improve your prototype if you had another week to work on it. What steps would you take to evaluate your results?

Note: Please refrain from using any existing NLP toolkit/library to solve this problem directly.

Good Luck!