Part 4:

What data structure did you finally use for vectors? What is the asymptotic memory usage of a vector? Of all your vectors? Is this memory usage reasonable and why?

What algorithm did you finally use for cosine similarity? What is its asymptotic running time? Is

this running time reasonable and why?

What algorithm did you finally use for the Top-J calculation? What is its asymptotic running time (might be in terms J, too)? Is this running time reasonable and why?

What improvements did you make from your original code to make it run faster? Give an example of your running time measurements before and after the changes. Describe asymptotic memory analysis, and/or profiling.

Part 5:  
Findings of Top-J comparing indexing one text to indexing two texts:

Finding on Top-J with three similarity measures

Part 6:

Plot of average distance vs. iteration; how many iterations to converge?