It turns out that shakespeare and beowulf had quotes like this "... 50 ..." which might give different output depending on the order you processed strings. So I have rerun the tests, removing all the double quotes from each corpus. I sent out the new reports. The new versions of each corpus have been uploaded to the test directory.

Can we get the the expected results posted to the test directory for these tests as well? Gmail does some weird newline text-wrapping thing that prevents the large results from being copied and pasted without changing.

You can get my report, which has the expected results:

wget beastie.cs.ua.edu/cs201/assign2/lusth@cs.ua.edu.rpt