

Homework #4

CS 5665, Fall 2016

100 points total [8% of your final grade]

Due: November 14, 2016 by 11:59pm

[no submission will be accepted after November 19, 2016 at 11:59pm]

Delivery: Submit via canvas

Overview

In this homework, you will write Map and Reduce functions to perform following two tasks:

Task 1: Word Count

A) Given the provided file ([Tolstoy's War and Peace](#)), create a complete count of each word that appears in the text. Which word appears the most?

B) Create a count of all the palindromes that occur in the text. Which palindrome occurs most often?

Task 2: Election Fraud

In this task your job is to investigate whether there was election fraud in 2008. You have 2006 and 2008 election data files: (i) [2006 data](#) file; and (ii) [2008 data](#) file. The files are of the format where each line is a vote in the election.

The format of the text file is:

VoterID \t CountyID \t PartyID

A) Which party won the election in 2008?

B) In 2006, which county was the most monolithic in the manner in which they voted? (i.e. which county came closest to voting 100% for a single party).

C) Studies have shown if a political party gains more than 50% in voting percentage from one election cycle to the next, then most likely fraud has occurred. (Example, if party A received 100 votes in 2006 in county B, then received 200 votes in 2008, fraud may have occurred). In which counties in 2008 did voter fraud likely occur?

D) From 2006 to 2008 how many voters changed which party they voted for? What is the most common type of change?

What to turn in:

- You should turn in a **PDF report** containing your answers, your **source codes** including Map and Reduce functions, and a **readme file** pointing which code is for what problem.

- This is an individual assignment, but you may discuss general strategies and approaches with other members of the class (refer to the syllabus for details of the homework collaboration policy). At the top of your report, please write the names of classmates you consulted and the nature of your discussion.