# Homework #4
## CS 5665, Fall 2016

## 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?

**Answer:** For counting the number of words in the input text file provided I have written map and reduce functions in a Java class named as WordCount.java. Then I executed the map and reduce functions for the input file and got the output. To find the word which appears the most in the file I opened the output file in excel and sorted the count column in the decreasing order and found that "the" word appeared the most which is 31700. I am attaching the map, reduce class, output file as object with this file.

WordCount.java        part-r-00000

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

**Answer:** For counting the number of palindromes in the input text file provided I have written map and reduce functions in a Java class named as PalindromeCounter.java. Then I executed the map and reduce functions for the input file and got the output. To find the palindrome word which appears the most in the file I opened the output file in excel and sorted the count column in the decreasing order and found that "a" word appeared the most (because the palindrome logic in my map considered single letters as palindromes) which is 10035. The palindrome string whose length was greater than 1 was "did" which appeared the most in this category for 1414 times. I am attaching the map, reduce class, output file as object with this file.

PalindromeCounter        part-r-00000
.java

## 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.

a) Which party won the election in 2008?

**Answer:** In this task, we need to find which party won the election in the input text file provided we need to find the count of votes each party got. I have written map and reduce functions in a Java class named as ElectionWinner2008.java. Then I executed the map and reduce functions for the input file and got the output. To find the party which won the election in the file I opened the output file and found that partyID "3" got most of the votes which is 12071. I am attaching the map, reduce class, output file as object with this file.

ElectionWinner200
8.java                part-r-00000

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).

**Answer:** In this task, we need to find which county became monolithic in the input text file provided we need to find the count of votes each party got in each county. I have written map and reduce functions in a Java class named as MonolithicCounty2006.java. Then I executed the map and reduce functions for the input file and got the output. To find the party which won the election in the file I opened the output file and found that County "333" became the monolithic for partyID "2" by giving most of the votes which is 45. I am attaching the map, reduce class, output file as object with this file.
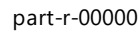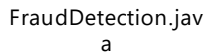
MonolithicCounty2
006.java             part-r-00000

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?

**Answer:** In this task, we need to find which county in 2008 voter fraud happened in the input text file provided we need to find the change of total votes each party got in each county in 2008 compared to 2006. If the increase in votes is more than 50% than there are chances of fraud. I have written map and reduce functions in a Java class named as FraudDetection.java. Then I executed the map and reduce functions for the input file and got the output. To find the party in each county which got number of votes increased by more than 50% in the file I opened the output file in excel and found following counties and the partyID which got votes increased more than 50% in the screenshot attached. I am attaching the map, reduce class, output file as object with this file.

| | A | B | C |
|---|---|---|---|
| 1 | 241 3 | 78 | |
| 2 | 178 3 | 72 | |
| 3 | 334 3 | 70 | |
| 4 | 332 3 | 65 | |
| 5 | 274 3 | 62 | |
| 6 | 390 3 | 61 | |
| 7 | 351 1 | 60 | |
| 8 | 332 1 | 52 | |
| 9 | 359 3 | 52 | |
| 10 | 424 3 | 52 | |
| 11 | 474 3 | 51 | |
| 12 | 324 3 | 50 | |
| 13 | 484 3 | 50 | |
| 14 | 232 3 | 47 | |
| 15 | 317 3 | 47 | |
| 16 | 148 3 | 45 | |

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

**Answer:** In this task, we need to find how many voters changed the party they voted for in 2006 in the input text file provided. I have written map and reduce functions in a Java class named as CountVotersChangedParty.java. Then I executed the map and reduce functions for the input file and got the output. To find the count of voters which changed party in 2008, I opened the output file in excel and found that 6297 voters changed their party.  For the second subtask in this task to find the most common type of change I implemented another mapper and reducer in the same java class which would only track the party changes and count the total number of changes for each change. I opened the output file and found that the most common party change is from party 2 to party 3 which is 1922.

I am attaching the map, reduce class, output files as object with this file.