

Assignment 2 - Code Review

Author: Shreysa Sharma

Reviewer: Jashangeet Singh

General Comments:

1. The code is quite efficient in that it only reads from the file twice, thereby improving the performance substantially.
2. The author has a class to read the entire file which enables checking for the neighbors across lines and splits, generalizing the neighbor mapping.
3. The code is a bit hard to follow due to some code repetitions and inefficient modularity.
4. The naming conventions are not always followed.

Minor Problems:

1. line 271: A Configuration object is named 'job'. Naming it 'job' is a bit confusing.
2. line 285: The property 'mapred.linerecordreader.maxlength' is deprecated in favor of 'mapreduce.input.linerecordreader.line.maxlength' as it says here (<https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/DeprecatedProperties.html>)
3. The code for lines 440:446 essentially performs the same function as lines 478:484 except to find the trailing neighbors instead of the leading ones. A separate method could be written to improve readability.