

Title : Code review for pdpmr-f17-deyb-a2-master

Created by: shabbirhussain

Created on: 10/02/17

Commit : 798c0189faa171da9662e7442ff8686282f226c8

General best practices

Code should be organized into subpackages like mapper, reducers, main. This helps anyone to understand the code structure easily. Plus it is easier to import from the package structure.

Eg. `package org.neu.pdpmr.a2.mappers;`

NeighborhoodScore.java:76

In the main function 3 things are done together, parameter parsing, executing job letter counter and executing job Kscoring. In general rule of thumb is to do one thing in one function. This function can be split into 4 functions.

1. `parseParams` doing parameter parsing
2. `letterScoringDriver`
3. `kScoringDriver`
4. `main` stitching them all.

CalculateScoreMapper.java:16

K has to be initialized once per mapper instance. So this method can be moved to setup

([https://hadoop.apache.org/docs/r2.8.0/api/org/apache/hadoop/mapreduce/Mapper.html#setup\(org.apache.hadoop.mapreduce.Mapper.Context\)](https://hadoop.apache.org/docs/r2.8.0/api/org/apache/hadoop/mapreduce/Mapper.html#setup(org.apache.hadoop.mapreduce.Mapper.Context)))) method of a mapper.

CalculateScoreMapper.java:26

Instead of creating an intermediate list and adding neighborhood words to it, then finally calculating score over it. A simpler approach will be to just calculate neighborhood scores for every word then and there only.

CountLetterReducer.java:10

`Reducer<Text, IntWritable, Text, IntWritable>` Instead of using `Text` and `IntWritable` as inputs you can use `IntWritable` (character `<->` int conversion is trivial) as key and `LongWritable` as values as count might exceed the integer range over the whole corpus.

CalculateScoreReducer.java:20

```
totalOccurance += 1; => totalOccurance ++;
```

Report

Readme.rmd

Minor typos:

```
out put -> output  pari -> pair
```

Makefile

- Starts with a capital 'M'.
- Please ensure tabs are appropriate in Makefiles. Currently compile, jar and run rules are not working due to tab issues. Proper syntax:

....

“” = space “” = tab