Lab Lecture #3.3

Word frequency in collection

```java
import java.io.IOException;
import java.util.HashMap;
import java.util.Map;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class WordFrequencyInCollection
{
   public static class NewMapper extends Mapper<LongWritable, Text, Text, Text>
   {
      public void map(LongWritable key, Text value, Context context)
         throws IOException, InterruptedException {
         String[] wordAndCounters = value.toString().split("\t");
         String[] wordAndDoc = wordAndCounters[0].split("@");
         context.write(new Text(wordAndDoc[0]),
                       new Text(wordAndDoc[1] + "=" + wordAndCounters[1]));
      }
   }

   public static class NewReducer extends Reducer<Text, Text, Text, Text>
   {
      public void reduce(Text key, Iterable<Text> values, Context context)
         throws IOException, InterruptedException {
         // total frequency of this word
         Map<String, String> tempMap = new HashMap<String, String>();
         int numberOfDocumentsInCorpusWhereKeyAppears = 0;
         for (Text val : values) {
            String[] docAndCounter = val.toString().split("=");
            tempMap.put(docAndCounter[0], docAndCounter[1]);
            numberOfDocumentsInCorpusWhereKeyAppears++;
         }
         for (String docKey: tempMap.keySet())
            context.write(new Text(key.toString() + "@" + docKey),
                          new Text(tempMap.get(docKey) + "/" +
                              numberOfDocumentsInCorpusWhereKeyAppears));
       }
   }

   public static void main(String[] args) throws Exception {
      Configuration conf = new Configuration();
      Job job = new Job(conf, "word frequency in collection");
      job.setJarByClass(WordFrequencyInCollection.class);

      job.setOutputKeyClass(Text.class);
      job.setOutputValueClass(Text.class);
      job.setMapperClass(NewMapper.class);
      job.setReducerClass(NewReducer.class);

      FileInputFormat.addInputPath(job, new Path(args[0]));
      FileOutputFormat.setOutputPath(job, new Path(args[1]));
      System.exit(job.waitForCompletion(true) ? 0 : 1);
   }
}
```