

Project 2 Grading

1.	CCR setup	(Should be able to run the Bigram Count program discussed in class)		10%
2.	Trigram Count Program	Mapper Code		20%
		Reducer Code		20%
3.	Project Report	a. Team Composition b. Connection to CCR Cluster c. Data Sets d. Design Issues e. Practical Experiences		20%
4.	Results and Discussion (A brief summary of your empirical results of your algorithm. Your summary should include answers to the following questions.)	a. How many unique trigrams were identified by your program?	10%	30%
		b. Does a change in the number of mappers and reducers affect the run time of your program?	15%	
		c. Did you have substantially long wait times before your job was processed?	5%	

Note:

- Please go through the Assignment document and the Project Report guidelines carefully.
- The Results and Discussion section should contain correct reasoning wherever necessary.
- Use of comments to explain the functions/statements in your code will be appreciated.
- There are no additional points for using regex or handling special characters and numbers in the input data.
- There are no additional points if you replicate the data files using Java or shell scripts.