1.

The author implements 5 approximate string matching algorithms to find a canonical form for each token in real Twitter data.

2

1)The structure and style of this report is scientific.

2)For the Related Work part, the author uses several references to introduce what people have done recently in spelling correction area, which is good.

3)For the Methods part, the author introduces how the system works briefly, which gives reader the blueprint of the whole project, which is good.

4)The author briefly introduces 5 algorithms instead of spending lots of space to introduce.

5)The author firstly analyzes the dataset and find that the dictionary is not suitable, and the explanation of this point is clear.

6)The author uses tables and figures to show the result, which is scientific.

7)The result is convincible, because it shows how many total returned predictions and how many correct predictions are, which is clear.

8)For analysis part, the author uses illustrative examples to prove his statement, which is convincible.

9)The author considers three noisy factors which influence the performance of algorithms, which is scientific.

10)The author considers future work, which is about updating the dictionary by adding some common abbreviation, homophones, and cyber words into dictionary, which is critical.

3

1)For the future work part, the author suggests the dictionary should add some abbreviations to make it more effective, which is not a difficult part. There are some libraries online, the author may add this into the report, which will make the report more scientific.

2)For the analysis part, the author spends more space in noisy factors instead of the difference of algorithms. That is because noisy factors greatly affect effectiveness so that all performances of different algorithms don’t work well.