Peer Review

Student 1

1. **Briefly summarise what the author has done**

The author has utilised Global Edit Distance, N-gram, Soundex for this project to find correct forms of misspellings. Precision, Recall, F-score are used to measure the results. The results are illustrated and the author has explained which algorithm is good for which type of spelling correction. The result shows N-gram is the best way for spelling correction to these datasets over the other two methods.

2. **Indicate what you think that the author has done well, and why**

1. The author utilises some figures and tables to illustrate the results, which is very clear and helpful for readers to understand the idea;
2. In the analysis part, the author put forward a lot of examples and provides potential solutions to deal with the problems, which is useful and insightful.
3. The report structure is organised, and methodologies are illustrated clearly.

3. **Indicate what you think could have been improved, and why**

1. In my understanding, the emphasis of this project is to gain knowledge based on the experiments results rather than comparing and finding a more accurate algorithm. According to the conclusion, it seems the author focuses more on the later one. The author has put forward many insightful ideas in the former part, such as the insufficient dictionary, but deeper analysis and illustration can be done as well.
2. When the author illustrates every method, the paragraph ends by suddenly saying whether the Precision and Recall increase or not. An explanation for these two terms and the reasons for fluctuation would help the reader to understand the report better. Or the author can put these sentences in part 4. In addition, the reason why the author chooses Recall and Precision should be justified. What do they for?
3. In my understanding, corpus means dictionary. The three files are datasets, not corpus.
4. Some conclusions are suddenly put forward (e.g. "...(N-gram's) Precision is expected to be higher than GED, while the Recall would be similar.").
5. It would be better to put some insight in the results part rather than the methodology part, such as the second paragraph of 3.1, which illustrates the reason why the correction fails is this dataset contains informal language.
6. The author has clearly illustrated the process for each methodology, but the output is not well defined. Until to end part of the results, I found the output is a list of candidates rather one word for each misspelling word.

Student 2

1. Briefly summarise what the author has done

1. Introduce the data set.
2. Using the method Global Edit Distance, N-grams and Soundex to introduce the analysis ways to solve the two kinds of problems, typing and pronunciation.
3. Introduce the evaluation metrics which contains three evaluations.
4. Analysis the outcome of the result and show the features of these three approaches.
5. Make a conclusion that N-gram is the best choice.

2. Indicate what you think that the author has done well, and why

1. The result samples shown in different methods provide a clear example to the reader and may improve the quality and the understanding to this report.
2. The background analysis at the Introduction part is clear, introducing the methods of the analysis and evaluation ways which showed explicitly.
3. The N-grams method contains the definition of the similarity which is better than only show the words introduction.
4. The author gives the hypothesis in all three methods and test them by the result, which is more clear for readers to read back.

3. Indicate what you think could have been improved, and why

1. The Soundex method shows an excessive deviation which may need to be discussed more of it. There may be some promotions to the code or the author may want to analysis the word from Twitter has some difference from other word when using the Soundex method. As far as I can see from this report it may shows this method is not useful.
2. The author may need to compare different N in N-gram method to show which N is better. Also, in the report, which N is used is not shown by the author.
3. It is better to show more figure or table to show the result analysis.
4. It is better to have the conclusion in future promotions.

Self-Reflection

1. **Briefly summarise what the author has done**

This report implements a misspell correction system for lexical normalisation based on three popular approximate string matching algorithms, the Global Edit Distance, the N-grams and the Soundex. The author gives an introduction of the lexical normalisation background and a brief summary of some published literature related to spelling correction and lexical normalisation. Then the author gives some details for approach implementation and gives some relevant hypothesis. Finally, the author analyses the results and gives some explanation about the reason related to the algorithms behind the results and proposes some potential approaches (e.g. a combination of two or more algorithms) for further improvement.

2. **Indicate what you think that the author has done well, and why**

In general, this is a very good report. The author compares the three algorithms and discusses their strength, weakness and makes some predictions about their precision and recall. For example, as the author mentioned, comparing with the Global Edit Distance algorithm, the N-grams is expected to have a higher precision and a similar recall. Also, the author gives some details that should be noted in the implementation. For example, if the comparison is based on similarity, the algorithm should return an empty list instead of the entire dictionary. In addition, the author introduces F-score as a more reliable evaluation indicator in the result analysis.

3. **Indicate what you think could have been improved, and why**

Although it is a good report in general, there are still a few things on this which could have been improved. The first one is digging some more information behind the evaluation indexes. For example, thinking about what kind of tokens that could be normalised through Soundex instead of GED (e.g. “plz” for “please”) because sometimes the precision and the recall may not tell us everything with details. In addition, as for the result of the Soundex, some analysis with more sample examples and more details would be better.