README

This project is using spelling correction to solve the problem of back–transliteration in the Persian language. Five approximate matching methods are involved in this project, which are Global Edit Distance, Global Edit Distance with replacement matrix,Local Edit Distance, N-gram and Soundex.

The programming language of this project is java.

The user processes of this project are shown below:

1. The system will print out “Please enter the root for train text: ” . The user will enter the root of train text in the console.
2. The system will print out “Please enter the root for name text: ” . The user will enter the root of name text in the console.
3. The system will print out the method choices for the user. And the user will enter an integer as the method number. If user enter a method number without definition, the system will print out “Wrong method number!”
4. After choosing the method the number, the system will run the method code and print out the precision and recall of the this method in the console(this operation may take 10 minutes or more). The format of the output is “ <method name>: precision is: <precision value> recall is: <recall value>”
5. Then the system will print out “Do you want to change the method?(Y/N)”. If the user want to use another method, the user has to enter “Y” and start from step 3 again. If not, enter “N”. The system will end and print out “end of the system” if user does not enter “Y”.

some other explanations:

1. the parameter i set in this system is through setCost method, i can change the parameter within the code, while the users may not change them.
2. the brief explanations for each method is above each method within the code.
3. the root I set for train text is “/Users/junwenz/Documents/project/kt/src/kt/train.txt” which the root of the train text in my own laptop.
4. the root I set for names text is “/Users/junwenz/Documents/project/kt/src/kt/names.txt” which the root of the names text in my own laptop.