

Spell Checker Write Up

Background

Searching is a powerful tool when navigating the internet, however, if a search is misspelled, determining what a user is looking for can be difficult. These mistakes are more obvious when a misspelled word does not match any word in the English language. Mistakes where the misspellings result in unintended words, on the other hand, are more difficult to detect. For example, if a user visits The Home Depot's website to search for kitchen faucets and enters "kitchen facet", the query contains no misspellings, but if the mistake is not identified the search results will not be what the user is looking for.

Problem Description

Your task is to develop an algorithm that will determine if a given search contains any misspellings.

Training Data

Training data will be provided in a file called checker_train.txt. Each line of the text will include a search phrase and a label of either correct or incorrect indicating if all the words within the phrase are spelled correctly. The search phrases and their correctness label will be in the following format -

Search Phrase|Correctness

Training data excerpt:

```
garden 4'' corrugated drain pipe|incorrect
galcier bay estates heritage bronze|incorrect
fluidmaster seals|correct
doors with mirrror|incorrect
mdf white board|correct
707222020068|correct
grey subway tile 3x9|correct
```

Testing

You will run your model with a set of test phrases and give us the predicted label (correctness) for each of the phrases in the form given below in "Output".

Input

The input to your classifier will be a test file called checker_input.txt. The file will have a search phrase on each line with no labels.

Output

The output of your classifier should be a text file in the format -

Search Phrase|Correctness

It should look like the training data file

Scoring

Scores will be determined by the F1 error of correctly detecting a misspelling