Documentation

💡 To extract the input and return a proper output, verifying every single element in the input string is necessary. Since the input is type of short string, the run time will not be long.

The Setup is as following:

1. Set up a database contains the brands we have or need to handle
    - Use a set to save time for lookup
2. Set up a database contains the unclear word close to brand
    - Use a hashmap(dictionary)
3. Build the database for country and state
4. use spelling check tool

The approach is as following:

1. Use functions to process the input
2. Build up uotput format
3. Convert input string to array/list without invalid symbols
4. Iterate the array and process every single element
    1. remove unnecessary sybols
    2. check if current element is in the brands database
    3. If no, check if it is in the unclear brands databae
    4. If no, try to correct the spelling and do 2 again
    5. If no, check if current element is a country

6. If no, check if current element is a state
7. if no, check if current element is a phone number

5. Return the proper output

Future work

Since the dataset is small, the cases are not complex.

For future work, the input check should be more strict.

1. Spelling correction should be more powerful.
2. The text replacement database(unclear brands) would be way larger.
3. Once we found a word(after correcting) that can be associated with a brand name, the replacement database is supposed to be updated by adding the original word maps to the correct brand name.
4. Expand the country and state database for more location usage.
5. To increase efficiency, an algorithm to verify words is necessary. In the current approach, if spell-check works bad, and the database does not contain a specific incorrect word, this approach will not perform well. On the other hand, if the database has a large size, the efficiency will terrible.