

Python Spell-Corrector

Ali Benazzouz

Mathis Goudjil

1/ Introduction :

The spell correction program developed uses the `symspellpy` library, a Python port of the SymSpell spell correction algorithm. SymSpell is a Symmetric Delete spelling correction algorithm that is known for its speed and accuracy. The algorithm precomputes possible spelling errors for every word in a dictionary to provide quick lookups and corrections.

2/ Methodology :

The program begins by initializing a `SymSpell` object with a maximum edit distance and prefix length. The maximum edit distance is the maximum number of edits that the algorithm considers when looking for corrections, while the prefix length is a parameter that affects the size of the created dictionary and lookup performance.

The program then loads a frequency dictionary into the `SymSpell` object. The frequency dictionary is a text file where each line contains a word and its frequency in a corpus, separated by a space. The frequency dictionary used in this program is `frequency_dictionary_en_82_765.txt`, which is included in the `symspellpy` library.

The spell correction function takes a string of text as input and uses the `lookup_compound` method of the `SymSpell` object to find corrections. The `lookup_compound` method can handle compound words and phrases, meaning it can find and correct misspelled words even if the input text is a phrase or a compound word.

The `lookup_compound` method returns a list of `SuggestItem` objects, each representing a possible correction. Each `SuggestItem` object contains the corrected term, the edit distance from the input term, and the frequency of the corrected term in the corpus. The program selects the corrected term from

each `SuggestItem` and joins them into a single string, which is the corrected version of the input text.

3/ Results :

The spell correction program can correct misspelled words in a text string, even if the text string contains compound words or phrases. The program selects the most likely correction based on the edit distance and the frequency of the corrected term in the corpus.

4/ Conclusion :

The spell correction program using `symspellpy` provides a fast and accurate way to correct spelling errors in text. It can handle compound words and phrases, making it suitable for correcting real-world text data.

However, the accuracy of the corrections depends on the quality and relevance of the frequency dictionary used. For specialized applications, a custom frequency dictionary may be needed.