SPELL CHECKING AND CORRECTING TECHNIQUES IN NLP

Abstract

Natural Language Processing's role of spelling correction is crucial. Search engines, sentiment analysis, text summarization, and other processes all use it. We strive to find and fix spelling issues in spelling correction, as the name implies. The correction of spelling errors in real-world NLP tasks helps models perform better because we frequently deal with data that contains typos. For users, unclear or inconsistent communication caused by misspelled words can be confusing. The automatic spelling suggestion provided by spell check improves readability and ensures communication clarity. Many packages are useful for spelling correction. Here, we have utilized TextBlob, Levenshtein and spellchecker packages for spelling correction and tested using input text.

Keywords: NLP; NLU; NLG; Spell Checker; Packages; Grammar; Library

Authors

Dr. P. Isakki Alias Devi
Associate Professor
Department of Computer Applications
Ayya Nadar Janaki Ammal College
Sivakasi, Tamilnadu, India.
harishvaidhyaa@gmail.com

Mrs. A. Agnes Saleema
Assistant Professor
Department of Computer Applications
Ayya Nadar Janaki Ammal College
Sivakasi, Tamilnadu, India.
agnessaleema09@gmail.com

I. INTRODUCTION

NLP stands for Natural Language Processing, is a subfield of computer science, humanities, and artificial intelligence. Machines can comprehend, analyse, manipulate, and interpret human languages because of technology. In order to execute tasks like translation, automatic summarization, Named Entity Recognition (NER), speech recognition, relationship extraction, and subject segmentation, it aids developers in organising knowledge. Figure 1 represents overview of NLP.
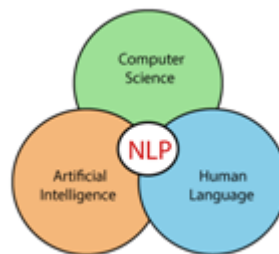

Figure 1: Overview of Natural Language Processing

Beginning in the early 1990s, NLP accelerated its growth and, particularly in English Grammar, attained high process accuracy. The introduction of an electronic text in 1990 also gave users a useful tool for training and testing natural language processing software. The availability of computers with quick CPUs and more memory may be one or more additional variables. The Internet was a key driver in the development of natural language processing.

NLP has several uses nowadays, including text reading, speech recognition, and machine translation. Combining all of these applications enables artificial intelligence to learn about the outside world. Let's use Amazon Alexa as an example. The human may ask Alexa a question and get a response from the robot.

1. **Advantages of NLP**
   - NLP enables users to ask inquiries about any topic and receive an immediate answer.
   - NLP gives precise answers to queries; it omits unnecessary or pointless information.
   - NLP enables computers to speak the languages of people.
   - It saves a lot of time.
   - The majority of businesses employ NLP to increase the accuracy and efficiency of documentation procedures as well as to extract information from huge databases.

2. **Disadvantages of NLP**
   - NLP might not display context.
   - NLP results are erratic
   - NLP might call for additional keystrokes.
   - NLP is designed for a single, narrow purpose since it cannot adapt to new domains and because it has a limited function.

II. COMPONENTS OF NLP

The two components of NLP are,

1. **Natural Language Understanding (NLU):** By removing metadata from text, such as concepts, entities, keywords, emotions, relations, and semantic roles, natural language understanding (NLU) enables computers to comprehend and interpret human language.

   NLU is primarily utilised in business applications to comprehend both verbal and written customer problems.

   The following activities are part of NLU.
   - The given input is mapped into a usable representation using this.
   - It is employed to examine many facets of the language.

2. **Natural Language Generation (NLG):** Natural Language Generation (NLG) serves as a translator, turning the computerised data into a representation in natural language. Text planning, Sentence planning, and Text Realisation are the three essential components. Figure 2 represents the differences between NLU and NLG.

Table 1: Differences between NLU and NLG

NLU	NLG
The act of reading and interpreting language is known as NLU.	The act of writing or generating language is known as NLG.
It produces non-linguistic outputs from natural language inputs.	From non-linguistic inputs, it creates natural language outputs.

III. APPLICATIONS OF NLP

The following are the applications of NLP:

1. **Smart Assistants:** In the modern world, a new smart device is introduced every day, making the world increasingly smarter. And not just machines are benefiting from this development. With the development of technology, we now have intelligent personal assistants like Siri, Alexa, and Cortana. They even respond to us in the same way when we speak to them in the same manner as we would.

   Natural Language Processing makes all of this feasible. By decomposing our language into its constituent elements of speech, root stem, and other linguistic qualities, it aids the computer system in comprehending it. It not only aids in language comprehension but also in processing meaning and emotions and responding in a human-like manner.

2. **Email Filters:** One of the simplest and earliest online applications of NLP is email filtering. Spam filters were the first to identify certain words or phrases that indicate a

spam message. But just like early NLP adaptations, filtering has improved. In Gmail's email classification, NLP is one of the more common and recent examples. Based on their content, the system determines whether emails fall into the primary, social, or promotional categories. This keeps your inbox, for all Gmail users, at a manageable size with critical, pertinent emails you want to analyse and respond to right away.

3. **Sentiment Analysis:** The term "opinion mining" is widely used to describe sentiment analysis. On the internet, it is used to evaluate the sender's attitude, behaviours, and emotional condition. The NLP (Natural Language Processing) and statistics used in this programme are used to give values to the text (positive, negative, or natural), as well as to determine the mood of the context (happy, sad, furious, etc.).

4. **Machine Translation:** Machine translation converts text or speech from one natural language to another. Grammar-wise, your Spanish assignment is a complete mess, which is one of the unmistakable signs that you plagiarised. Translation services used to ignore the fact that many languages have different sentence structures and cannot be translated directly. But they have made great progress. Online translators are able to translate languages more correctly and provide grammatically sound outcomes because to NLP. When attempting to converse with someone who speaks a different language, this is really helpful when attempting to communicate with someone who speaks a foreign language. Example: Google Translator

5. **Spelling Correction:** For spelling correction, Microsoft Corporation offers word processing programmes like MS-word and PowerPoint. Millions of texts are used to train the NLP algorithm so that it can recognise the proper structure. Because of this, it may offer a better synonym, the proper verb tense, or a clearer sentence structure than what you have typed. Grammarly, WhiteSmoke, ProWritingAid, and other popular NLP-based grammar checkers are very useful for spelling correction.

6. **Speech Recognition:** The transcription of spoken words into text is called speech recognition. Applications for it include voice user interface, voice biometrics, Microsoft Word dictation, mobile, home automation, video recovery, and more.

7. **Chatbot:** Everything has gone digital as a result of the advancement of technology, including education, buying, booking travel arrangements, and customer service. The chatbot responds promptly and precisely. These chatbots are equipped with conversational capabilities thanks to NLP, enabling more accurate consumer responses than simple one-word answers.

   Chatbots are also useful in situations when human labour is scarce or not available constantly. NLP-based chatbots also include emotional intelligence, which enables them to recognise and successfully address the emotional needs of their users.

8. **Information Extraction:** One of the most crucial uses for NLP is information extraction. It is used to extract structured data from machine-readable documents that are unstructured or semi-structured.

IV. AI GRAMMAR AND SPELL CHECKER

An AI Grammar and Spell Checker is a program that is used to automatically detect and fix grammatical and spelling errors in written text. It works by reading the text, dissecting it, and then comparing it to a massive library of grammar and spelling rules as well as common errors.

Using a basic spell checker, the following tasks are completed: The text is examined, and the words that are present are extracted. A dictionary is used as a database of known correctly spelt words after which each word is evaluated. Another stage is to discuss morphology in the context of a particular language.

V. PACKAGES FOR SPELL CHECKER AND CORRECTOR

TextBlob, Levenshtein and spellchecker packages are useful for spelling correction.

1. **TextBlob:** A powerful Python NLP package called TextBlob offers a simple user interface for processing textual data. It contains a spell-checking facility that you may use to fix typos in text. To make spelling fixes, TextBlob combines edit distance measurements with word frequency data.

We may create a Python spell checker and corrector utilising NLP by using the TextBlob package, which provides a straightforward mechanism to perform spelling correction. You can use a spell checker and corrector by doing the following:

```
from textblob import TextBlob

def correct_spell_checker(text):
    blob = TextBlob(text)
    corrected_text = blob.correct()
    return corrected_text

# Example usage
input_text = "The dataa minng is the proces of sortig throgh larg data sets"
corrected_text = correct_spell_checker(input_text)
print("*****************************************************************")
print("\t\tSpell Checker & Corrector using TextBlob")
print("*****************************************************************")
print(input_text)
print(corrected_text)
print("*****************************************************************")
```

```
*****************************************************************
              Spell Checker & Corrector using TextBlob
*****************************************************************
The dataa minng is the proces of sortig throgh larg data sets
The data mining is the process of sorting through large data sets
*****************************************************************
```

Figure 2: Using TextBlob for Specific Text

Figure 2 represents the Python code using TextBlob for spell checking and correction. You can now read "The data mining is the process of sorting through large data sets" in place of the previously misspelt words in the line "The dataa minng is the

proces of sortig throgh larg data sets". Please be advised that the accuracy of the spelling correction will depend on the underlying language models used by TextBlob. If you require a more complex spell checker and corrector, you can look at additional NLP libraries and techniques, such as using personalised language models or external spelling correction datasets.

```
[2]  from textblob import TextBlob

     def spell_checker_corrector(text):
         blob = TextBlob(text)
         corrected_text = blob.correct()
         return str(corrected_text)

     # Example usage
     input_text = input(" Enter the text")
     corrected_text = spell_checker_corrector(input_text)
     print("Original text:", input_text)
     print("Corrected text:", corrected_text)


      Enter the textHello, hwo are yuo?
     Original text: Hello, hwo are yuo?
     Corrected text: Hello, who are you?
```

Figure 3: Using TextBlob for User Text

Figure 3 represents the Python code using TextBlob for spell checking and correction for user input. The original text is "Hello, hwo are you?" and the corrected text is "Hello, who are you?"

2. **Levenshtein:** Utilising the Levenshtein package, you're putting into practise a simple spelling correction function based on the Levenshtein distance algorithm. The function receives a list of words with the correct spellings and one misspelt word. It returns the word with the shortest Levenshtein distance from each of the corrected words as the corrected term. Levenshtein is imported using the Pip install command. The correct_spelling function, which you create, accepts two inputs: a vocabulary list and a misspelt word. The Levenshtein distance between each correctly spelt word in the dictionary and the incorrect term is calculated using the function. Following that, the method returns the updated word with the shortest distance.

Before utilising the function to repair a term that has been entered wrongly, you ask the user to compare it to the vocabulary list. The term that was repaired is then printed in the output. Overall, the spelling correction approach for the Levenshtein distance is simple but efficient. The performance of a spelling correction system can be improved by employing more advanced approaches, such as language models or external dictionaries, instead of the Levenshtein distance, which might not always provide the most accurate results. Figure 4 represents the Python code using Levenshtein for spell checking and correction.

```
✓ [30] !pip install Levenshtein

    def correct_spelling(word, vocabulary):
        min_distance = float('inf')
        corrected_word = word

        for correct_word in vocabulary:
            distance = Levenshtein.distance(word, correct_word)
            if distance < min_distance:
                min_distance = distance
                corrected_word = correct_word

        return corrected_word

    # Example usage
    vocabulary = ["apple", "banana", "orange", "grape"]
    misspelled_word = input("Enter the word ")
    corrected_word = correct_spelling(misspelled_word, vocabulary)
    print(corrected_word)

Requirement already satisfied: Levenshtein in /usr/local/lib/python3.10/dist-packages (0.21.1)
Requirement already satisfied: rapidfuzz<4.0.0,>=2.3.0 in /usr/local/lib/python3.10/dist-packages (from Levenshtein) (3.1.2)
Enter the word orng
orange
```

Figure 4: Spell Checker and Corrector using Levenshtein

3. **Spellchecker:** A professional Python library for spell checking and correction is the spellchecker package. It offers a simple method to find and fix spelling mistakes in text. This Python programme uses the spellchecker module to determine whether a word entered by the user is spelt correctly. The code is broken down as follows: The SpellChecker class from the spellchecker library must first be imported. Then, to store misspelled words, construct an empty set called misspelled_words. When determining whether a word is misspelled, this set will be used for quicker lookup. After that, assign a new instance of the SpellChecker class to the variable spell. Figure 5 represents the Python code using spellchecker package for spell checking and correction.

```
✓ [3] pip install pyspellchecker

    Collecting pyspellchecker
      Downloading pyspellchecker-0.7.2-py3-none-any.whl (3.4 MB)
                  ━━━━━━━━━━━━━━━━━━━━━━━ 3.4/3.4 MB 15.3 MB/s eta 0:00:00
    Installing collected packages: pyspellchecker
    Successfully installed pyspellchecker-0.7.2

✓ [6] #import spellchecker library
    from spellchecker import SpellChecker
     #create a variable spell and instance as spellchecker()
    misspelled_words = set()
    spell=SpellChecker()
    '''Create a while loop under this loop you need to create a variable called a
    word and make this variable that takes the real-time inputs from the user.'''
    w=input("Enter any word of your choice:")
    w=w.lower()
    print(w)
    '''if the word that presents in the spellchecker dictionary, It
    will print "you spelled correctly" Else you need to find the best spelling for that word'''
    if w in spell.word_frequency:
        print("'{}' is spelled correctly!".format(w))
    else:
        correctwords=spell.correction(w)
        print("The best suggestion for '{}' is '{}'".format(w,correctwords))
        # add the misspelled word to the set for faster lookup
    misspelled_words.add(w)

    Enter any word of your choice:kyboad
    kyboad
    The best suggestion for 'kyboad' is 'keyboard'
```

Figure 5: Spell Checker and Corrector using Spellchecker

VI. CONCLUSION

A tool that is used to automatically find and correct grammatical and spelling problems in written text is known as an AI Grammar and Spell Checker. It operates by reading the text, analyzing it, and then contrasting it with a vast database of grammar and spelling rules as well as typical mistakes.

The following tasks are carried out using a rudimentary spell checker: The words that are present in the text are extracted after careful examination. Each word is then examined using a dictionary as a database of recognized correctly spelled words. A subsequent step is to talk about morphology in the context of a certain language. The packages like TextBlob, Levenshtein and spellchecker can be useful. The spell checker and corrector coding have been written in Python and tested with uncorrected, misspelled text or sentences and get correct outputs.

REFERENCES

[1] Bañón, M.; Chichirau, M.; Esplà-Gomis, M.; Forcada, M.L.; Galiano-Jiménez, A.; García-Romero, C.; Kuzman, T.; Ljubešic´, N.; van Noord, R.; Pla Sempere, L. Croatian Web Corpus MaCoCu-hr 2.0. Slovenian Language Resource Repository CLARIN.SI, ISSN 2820-4042, 2023.

[2] Ding, Y.and Luo, X., "Applications of natural language processing in construction", Elsevier, Volume 136, 2022.

[3] Gledec, G.; Šoic´, R.; Dembitz, Š. Dynamic N-Gram System Based on an Online Croatian Spellchecking Service. IEEE Access 2019,7, pp.149988–149995, 2019.

[4] Jayanthi, S. M., Pruthi, D. and Neubig, G., "NeuSpell: A neural spelling correction toolkit," in Proc. the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pp. 158–164, 2020.

[5] Khurana, D., Koli, A. and Khatter, K., "Natural language processing: state of the art, current trends and challenges", Springer, 2022.

[6] Linna, N., Charles, E. and Kahn, K., "Applications of natural language processing in radiology: A systematic review", International Journal of Medical Informatics, Volume 163, 2022.

[7] Kinaci, A. C., "Spelling correction using recurrent neural networks and character level N-gram," in Proc. 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), Malatya, Turkey, pp. 1–4, 2018.

[8] Mijanur Rahman, M. D., Hasan Mahmud, Razia Sultana Rupa, and Mahnuma Rahman Rinty. A ro- bust bangla spell checker for search engine. In 2021 6th International Conference on Communica- tion and Electronics Systems (ICCES), pp:1329– 1334, 2021.

[9] Pancholi, A., Jain, R. and Patel, M., "A Review Paper on Applications of Natural Language Processing-Transformation from Data-Driven to Intelligence-Driven", IJERT, Volume 10, Issue 05, 2021.

[10] Sooraj, S., Manjusha, K., Kumar, M. and Soman, K., "Deep learning based spell checker for Malayalam language," Journal of Intelligent and Fuzzy Systems, vol. 34, pp. 1427–1434, 2018.

[11] Ulcˇar, M.; Supej, A.; Robnik-Šikonja, M.; Pollak, S. Slovene and Croatian Word Embeddings in Terms of Gender Occupational Analogies. Slov. 2.0 Empir. Appl. Interdiscip. Res. 9, pp:26–59, 2021.