Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Natural Language Processing

DIGITAL ASSIGNMENT I

<u>Submitted by:</u>

Name   : Shashank V

Reg No   : 15BCE0042

Slot   : E1

Course Professor : Arivoli A

Synopsis:

Question: Introduction of Real-Time application of NLP in Spell and Grammar Checker and tools used for the same.

1.1) What is Natural Language Processing?

Natural Language Processing, or NLP for short, is broadly defined as the automatic manipulation of natural language, like speech and text, by software. The study of natural language processing has been around for more than 50 years and grew out of the field of linguistics with the rise of computers.

1.2) Evolution of Natural Language Processing:

- While natural language processing isn't a new science, the technology is rapidly advancing thanks to an increased interest in human-to-machine communications, plus an availability of big data, powerful computing and enhanced algorithms.

- As a human, you may speak and write in English, Spanish or Chinese. But a computer's native language – known as machine code or machine language – is largely incomprehensible to most people. At your device's lowest levels, communication occurs not with words but through millions of zeros and ones that produce logical actions.

- Indeed, programmers used punch cards to communicate with the first computers 70 years ago. This manual and arduous process was understood by a relatively small number of people. Now you can say, "Alexa, I like this song," and a device playing music in your home will lower the volume and reply, "OK. Rating saved," in a humanlike voice. Then it adapts its algorithm to play that song – and others like it – the next time you listen to that music station.

- Natural language processing helps computers communicate with humans in their own language and scales other language-related tasks. For example, NLP makes it possible for computers to read text, hear speech, interpret it, measure sentiment and determine which parts are important.

1.3) How does NLP work?

Natural language processing includes many different techniques for interpreting human language, ranging from statistical and machine learning methods to rules-based and algorithmic approaches. We need a broad array of approaches because the text- and voice-based data varies widely, as do the practical applications.

Basic NLP tasks include tokenization and parsing, lemmatization/stemming, part-of-speech tagging, language detection and identification of semantic relationships. If you ever diagramed sentences in grade school, you've done these tasks manually before.

In general terms, NLP tasks break down language into shorter, elemental pieces, try to understand relationships between the pieces and explore how the pieces work together to create meaning.

2.1) NLP methods and applications

NLP text analytics:

Natural language processing goes hand in hand with text analytics, which counts, groups and categorizes words to extract structure and meaning from large volumes of content. Text analytics is used to explore textual content and derive new variables from raw text that may be visualized, filtered, or used as inputs to predictive models or other statistical methods.

NLP and text analytics are used together for many applications, including:

Investigative discovery. Identify patterns and clues in emails or written reports to help detect and solve crimes.
Subject-matter expertise. Classify content into meaningful topics so you can take action and discover trends.
Social media analytics. Track awareness and sentiment about specific topics and identify key influencers.

Everyday life NLP examples:
There are many common and practical applications of NLP in our everyday lives. Beyond conversing with virtual assistants like Alexa or Siri, here are a few more examples:

Emails in your spam folder and noticed similarities in the subject lines. You're seeing Bayesian spam filtering, a statistical NLP technique that compares the words in spam to valid emails to identify junk mail.

Missed a phone call and read the automatic transcript of the voicemail in your email inbox or smartphone app. That's speech-to-text conversion, an NLP capability.

Navigated a website by using its built-in search bar, or by selecting suggested topic, entity or category tags. Then you've used NLP methods for search, topic modeling, entity extraction and content categorization.

2.2) Spell and Grammar Checker:

Spell checker is a program or feature of program which is used to recognize the words which are misspelled and inform the user about those misspelled words. It depends on the feature of spell checker either to automatically correct the misspelled word or ask the user to choose the correct word from the suggestions which are available for that particular misspelled word. Spell checker may be an application which has capability of working on a large part of text or as an element of bigger application such as text editor, email, blog writing, keyword searching.
The main steps accomplished by spell checker are:-

1. Receives the sentence or word as input.
2. Divide the sentence into words and also process the word to make it suitable for transferring it to the next step.
3. The word is searched in the lexicon.
4. If the word is present in the lexicon, next word is processed.
5. In case the word does not exist in the lexicon, then its closest matching words are searched for in the lexicon and then given to the user as suggestions.

Grammar checker is an application or an element of application that recognizes written text for grammatical errors and then corrects those errors. Almost all the grammar checkers are executed as an element of a bigger application, such as word document, email but they are also accessible as an application which may be standalone that can be prompted from within different types of programs that work with edible text. Natural language processing is mainly used for developing a grammar checker. An example of a software program that includes its own grammar checker is **Microsoft Word**.

Applications of grammar and spell check are available for various languages like English, Chinese and Tamil. Taking the example of Sinhala, the national language of Sri Lanka, the users when type a word in the language can be automatically corrected using spelling and grammar correction tools and in fact provide real time suggestions to the end users when they type a sentence.
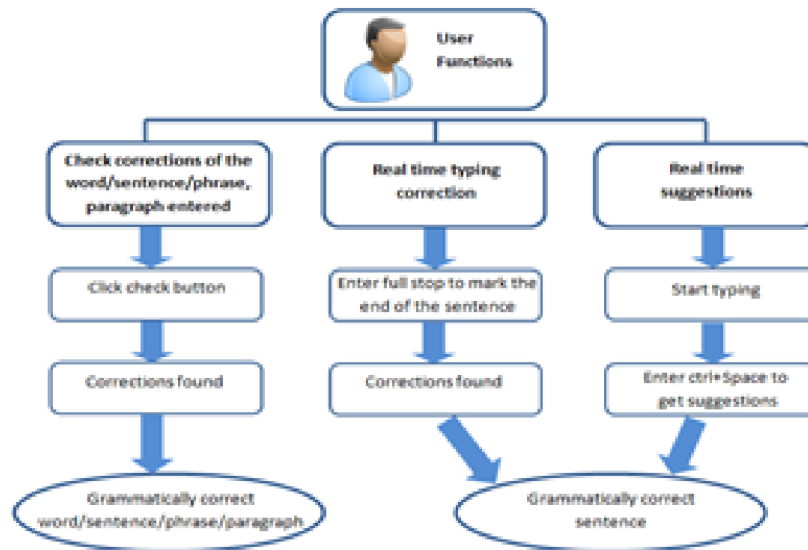
A comprehensive overview of how the users functions are passed through various functions and finally corrected grammatically as a word/sentence/phrase/paragraph. The implicit methods implemented in the checking are as follows:

According to the grammar identified during the analysis phase by the lexicon, a new lexicon will be created to map the words to create a correct sentence. According to the updated lexicon the system will gradually develop to suggest different solutions to correct a grammatical error. Not only that, tried out another ways to do this work to analyze the effectiveness way of doing grammar correction. Using neural network system, tried out two different ways of implementations. Class based text classification and mathematical model text classification. In class based text classification, sentences separated into the classes to train. In that case, need more data set with more classes. This text classification method already implemented for English It trained as greeting, good bye and sandwich kind of categories. Data set implemented as category wise. Then tried out to a new sentence, which matching with one of existing categories.

2.3) Tools for Spell and Grammar Checker:

The spelling checker tools used in present websites are:

a) Grammarly: The Grammarly Keyboard helps you write mistake-free and with ease in any app you use on your mobile device. It provides hundreds of checks and features, along with seamless integration.

b) Ginger: Ginger corrects your typos, phonetic mistakes, severe spelling mistakes, and misused words based on the context of a sentence. With just a single click, you can correct all your spelling mistakes to make your content smooth.

c) Spell Checker: A spell checker is an application program that flags words in a document that may not be spelled correctly.