

Natural Language Processing Semantical and Syntactical Analysis for English

Gayatri P Patil¹

Department of Computer Science
Rani Channamma University,
Belagavi-591156, India
E-mail: gayatripatil865@gmail.com

Mallamma V Reddy²

Department of Computer Science
Rani Channamma University,
Belagavi-591156, India
E-mail: mallammantreddy@gmail.com

Suman P Gouda³

Department of Computer Science
Rani Channamma University
Belagavi-591156, India
E-mail: sumangouda123@gmail.com

Nishali R Nandarge⁴

Department of Computer Science
Rani Channamma University,
Belagavi-591156, India
E-mail: nishalinandarge@gmail.com

Abstract—Natural language is to facilitate the user to exchange the ideas among people. These ideas converge to form the "meaning" of an utterance or text in the form of a series of sentences. The meaning of sentences describes as semantics. The input/output of a NLP can be a written text or a speech. There are two major components of natural language processing, namely: natural language understanding which describes mapping of given input in the natural language into a useful representation and the other is natural language generation which produce natural language as output on basis of input data as text. This paper deals with natural language understanding mainly on semantics

Keywords- compound words, Morphology, NLP, pragmatics, Semantics, syntactic

I. INTRODUCTION

The semantical study of any language is comprises of syntax, semantics and pragmatics [1], morphology. The term syntax is a part of grammar which tells the rules that governs the structure of sentences. It also concerns with how words grouped and connected to each other in sentence. Semantics it is a branch of linguistics, which aims to study the meaning of a language. It covers most complex tasks like: finding synonyms, word sense disambiguation, translating from one natural language to another. To solve any semantical problem morphological and syntactical analysis must be done.

Grammatical (syntactically valid) form of English is "SUBJECT VERB OBJECT" and for natural language like Kannada is "SUBJECT OBJECT VERB" does not imply sensible (semantically valid). The grammatical sentence are shown in example 1 and 2,

Example 1:

"Dog eats Cat" } –statement 1

ನಾಯಿ ಬೆಕ್ಕನು ತಿನ್ನುತ್ತದೆ.

"Cat eats Dog" } - statement 2

ಬೆಕ್ಕು ನಾಯಿಯನ್ನು ತಿನ್ನುತ್ತದೆ.

Both are syntactically valid that is it contains subject object verb but the statement 2 is semantically wrong that the general knowledge about the world, which each language user must know about the other's beliefs and goals. Hence here the general knowledge is "dog eats cat" but "cat does not eat dog" and the other example is

Example 2:

"You have a red light"

ಕೆಂಪು ಬತ್ತಿಯನ್ನು ಉರಿಸಿ. [Kempu battiyannu urisi]

The sentence in example 2 is grammatically ok (subject verb object) in English and (subject object verb) in Kannada, but makes no sense. The sentences are used in different situations and how that affects the interpretation of sentence is described by pragmatics [2]. The sentence "You have a red light" is ambiguous. Without knowing the background, the identity of the speaker, and his or her intention, it is difficult to understand the meaning with confidence. For example the sentence "You have a red light", it could mean that:

- You have red ambient light.
- You have red light while driving your car.
- You cannot go ahead with your project.

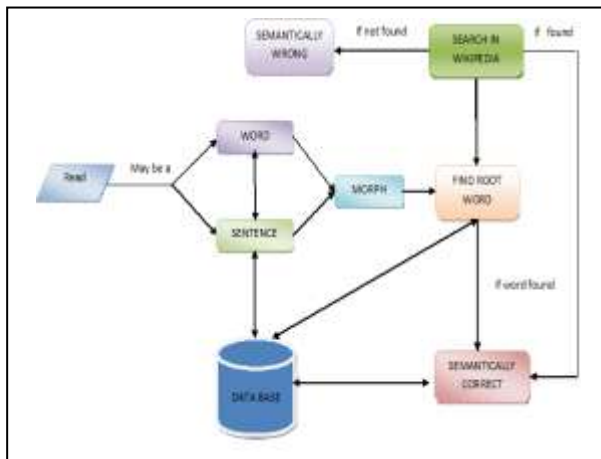


Figure 1 System Architecture for semantic analysis

The system accepts input in form of character, word or sentence. If the input is character it will check for valid ASCII alphabets, numerals and special characters then verifies its semantics. If the input is root word it matches to the database means it's semantically right or if the input is compound word morphing will be applied to get a root word for its semantic Correctness for example- working where the root word is 'work' and 'ing' is suffix. If the input is a sentence we have to check for character, word semantic and sentence semantic which should be in the form of Subject Verb Object for English sentence which exist in world discourse knowledge. If root word is not present in the database then search the word in Wikipedia and display the appropriate result.

II. SEMANTICAL ANALYSIS

Semantical analysis is the part of natural language processing, Semantical analysis includes the following terminologies

A. Syntax

It is the study of structural relationship between the words. [3]

For example: Man threw ball.

The given above example is in English language, which consist of Subject + Verb + Object.

I.e. Man (Subject) + threw (Verb) + ball (Object).

B. Semantics

It is a branch of linguistics which deals with understanding the meaning of language.

For example 1: Some see the glass half empty and others see the glass half full.

It has the real world knowledge hence it is semantically correct sentence.

Words without Meanings where Advertisers use certain words, some without any real meaning at all, to convey certain impressions

For example 2: Do you have tired blood? [4]

C. Compound word

It is the word which is followed by either suffix or a prefix.

For example: playing, where the root word is 'play' and suffix is 'ing'.

D. Morphology

It is the branch of linguistics, which deals with study of word structure. How they are formed and their relationship to other words in the sentence. It is the process of analyzing the structure and parts of words, such as stems, root words, suffixes and prefixes [5].

For example – singing, this is a compound word where root word is 'sing' and suffix is 'ing'.

E. Pragmatics

It is the study of different aspects of language use in particular situations, which deals with the ways in which speaker and listener use and interprets words and utterances [6].

For example: she is taking turns in conversations

III. METHODOLOGY

Natural Language Processing is a system which takes input as string of words and produce structured representation as output by capturing the meaning of those strings. This paper carries Semantical analysis for the Natural Language such as English which is a global language. The algorithm 1.1 is designed to produce an experimental system in Semantical analysis for English by providing input as characters, words and sentences. The output of the system checks the Semantical correctness of the given input (character, word or sentence).

Input: English word or sentence

Output: semantically analyzed result

Algorithm 1.1

Step 1: Read the input

Step 2: Check whether the entered input is character, or a word, or a sentence.

Step 3: If the input is a character then check for its valid ASCII alphabets, numbers and special character. Go to step 6

Else

Step 4: If the input is a word then check the Semantical availability of the word in database.

a. If the word is found in the database, go to step 6.

Else

- b. If the word is compound word then separate the suffix from the root by applying morphological rule. Then re-initialize the word without suffix.
- c. If the input word is not found in the database, check the Semantical availability of the word in Wikipedia.
- d. Repeat step 4 until the database finds the root word.

Step 5: If the input is a sentence split the sentence by

- a. A character then goes to step 3
- b. Word semantic then go to step 4

Then verify the sentence for semantic which is in the form of Subject Verb Object. Go to step 6

Step 6: Display the entered input is semantically right.

Else

Step 7: Display the entered input is semantically wrong

IV. IMPLEMENTATION RESULTS

The Semantical analysis for English language is built by using HTML [7] as front end and WAMP server [8/] for database. The system is implemented for characters which includes the numbers, characters and special characters and compound words where morphological rules are applied to get root word and then checks it for Semantical correctness. The sample results are shown in Figure 2 to 6

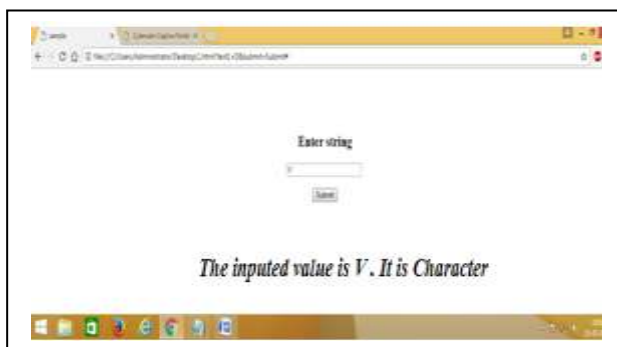


Figure 2 Semantical analysis for character

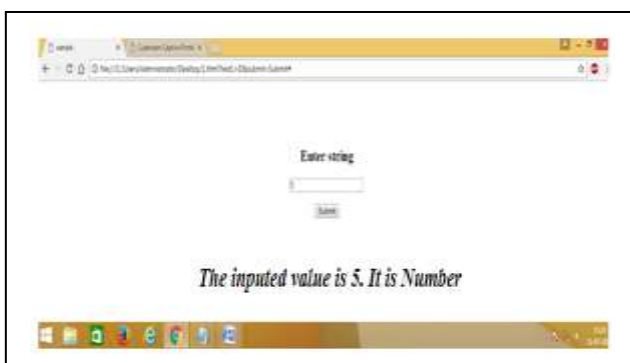


Figure 3 Semantical analysis for Number

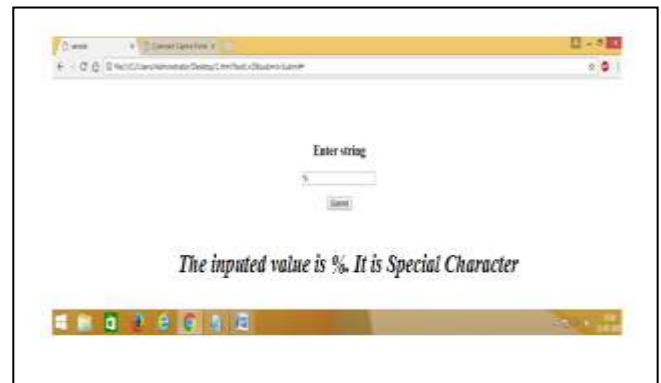


Figure 4 Semantical analysis for Special characte



Figure 5 Semantical analysis for Compound Words

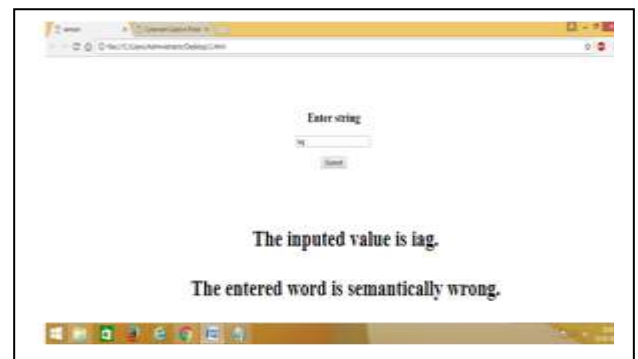


Figure 6 Semantical analysis for non existing word

V. CONCLUSION AND FUTURE WORK

This paper presents an approach of semantical analysis for English language which applies on characters and compound words; it is useful in finding synonyms, word sense disambiguation, translating from one natural language to another and populating base knowledge. The work will be continued for checking the semantical correctness of sentences and further same will be applied for natural languages.

REFERENCES

- [1] http://www.nutrained.com/language/syntax_semantics.htm
- [2] <http://www.jguru.com/faq/view.jsp?EID=81>.

- [3] <https://www.thoughtco.com/subject-verb-object-1692011>
- [4] <http://examples.yourdictionary.com/examples-of-semantic.html>
- [5] https://www.cs.bham.ac.uk/~pjh/sem1a5/pt2/pt2_intro_morphology.html
- [6] <http://study.com/academy/lesson/what-is-pragmatics-definition-examples.html>
- [7] https://www.w3schools.com/html/html_intro.asp:
- [8] <http://www.wampserver.com/en/>.