EasyChair Preprint

№ 7676

A Survey Paper for Parsing of English Sentences Using Context-Free Grammar

Chetanya Rathi, Harsh Dhabekar, Harshit Mundhra,
Pushkar Helge and Kartik Rajput

March 29, 2022

A survey paper for Parsing of English sentences using context-free grammar

Chetanya Rathi
Artificial Intelligence and Data Science
Vishwakarma Institute of Technology
Pune, India
chetanya.rathi20@vit.edu

Harsh Dhabekar
Artificial Intelligence and Data Science
Vishwakarma Institute of Technology
Pune, India
harsh.dhabekar20@vit.edu

Harshit Mundhra
Artificial Intelligence and Data Science
Vishwakarma Institute of Technology
Pune, India
harshit.mundhra20@vit.edu

Pushkar Helge
Artificial Intelligence and Data Science
Vishwakarma Institute of Technology
Pune, India
pushkar.helge20@vit.edu

Kartik Rajput
Artificial Intelligence and Data Science
Vishwakarma Institute of Technology
Pune, India
kartik.rajput20@vit.edu

***Abstract*— The objective of parsing is to transform a natural language sentence it in to a standard order. and in a same way a sentence is tokenized with an appropriate format. There are certain English grammar evaluation rules and the parsing approach which is to be followed for the proper formation of a particular sentence syntactically and semantically using the parsing approach. A sentence in English language is the main element in the semantic parser, which creates a parse tree with the help of applying semantic dating technique to a number of phrases. A parser divides a token into smaller components by applying sets of guidelines that characterize and a series of the tokens to determine its structure of the language, which specified by grammar. The illustration provides easy records on grammatical connections, which can simply know and put into practice with those who have no prior knowledge of the language, such as those who need to obtain textual family members. The semantic family members represent the relationships of a number of the words in the sentence. We advocate utilizing our parser to acquire the tagged sets as well as a context-free layout grammatical representation for the source form. All pronouns, adverbs, singular, plural, nouns, verbs, people, adjectives, tenses and other words are kept in a database.**

Keywords*— *Artificial Intelligence, Bottom-up Parser, Context-Free-Grammar, English Grammar, Python, Parse Table, Semantic Parser, Top-down Parser.

I. INTRODUCTION

Natural Language Processing is subject of pc technological know-how, linguistics and AI, involved with the interaction between human and computer. It's far from ambition of test and alertness that analyses how computer systems may be used to manage speech or natural language text. Although, that enables us to interpret this phrase verbally or actually in gadget media is important, which plays as an interpreter and compiler of a device. Semantic parser is the name for this intermediary between humans and machines. It is a procedure that depicts the natural language to a identified format while keeping the unique meaning. The parser performs semantic reading on the textual material in order to do such semantic operations.

The analysis is carried out using a large database and a set of elements that specify the sentence construction of the phrase as well as the right interpretation of the sentence. Normally, a semantically parser is concerned with the grammar rules and word family members of a phrase. Moreover, it determines the duties of the phrases in the text Sentence using knowledge assets and grammatical regulations. Many compositions have already been completed for the English language, [1] and many of them confirmed astounding obligations. almost 1.35 billion people inside the international community speak this language. Almost 1.35 billion people inside the international community speak this language. As a result, a method such as semantically parsing of English sentences, which might be implemented to meaningfully maps the English language sentence, may be extremely valuable.

In addition to determining the grammatical structure, an English semantic parser may also identify the meaning of the phrase. We'll need an English semantic parser to begin understanding the English language's semantics. In a global commercial enterprise, government and science related subjects and amusement, all web sites are relaxed in the English language. Moreover, the effect of English computing isn't always restricted to socio-economic destiny. It will also discuss English language unity in the global context. The researcher's ultimate goal is to improve a

computer translation device for the English , which still in its early phases of development and has limitations. In this paper, we give an English Semantic parser which includes a reading method for language sentences to enter each semantically and syntactically. A set of recognized semantic criteria is used to ensure the phrase's sentence rules as well as semantic meaning.

The method is broken down into 3 steps: first is to, find out what the terms in the statement signify, then validate grammatical form of sentence using the set of grammatical rules, and the last is , justify significant linkages between phrases to confirm the a relevant phrase. The major purpose of the English parser is to develop a method for analysing, understanding, and creating human languages in order for, in end It'll be able to speak with the computers as if it were talking to someone else. Because textual material may store knowledge at a variety of granularities, ranging from low-level token–based representation to complex higher–level logical representation across the document collections, this study aims to paint at the appropriate level of analysis for applications in question. These investigations work by implementing an unique new way for parsing English sentences.

II. Literature Review

The authors of paper [1] have proposed a Bangla grammar parser, they have gone through various steps to achieve this functionality. First step they followed was to store the words in the xml format which is nowadays widely used. In the extensible language format, they did design a dictionary kind of structure to store its English meaning inside the Bangla tag angular brackets. Second step which they followed was to develop a predefined set of words for Bangla they did keep a particular syntax for some specified kinds of differentiating features to follow a defined representation of the meaningful sentence obtained by composing single or multiple phrases. In the third step they started identifying the actual elements for getting meaningful context using context free grammar they defined some development rules for the same and also used the method of left factoring. The final step was designing a parser by defining two important functions which determined how the sentences would enter and helped making grammar table for respective sentences.

Author in the paper [2] have discussed about visualizing derivations in grammar, author has proposed a system for the same by using programming languages like python, stream lit which is a popular framework. They are accepting development rule in string format from the user which are further saved into a text file using some basic python operations. Further the user can again input some string and the system will check if the entered phrase is a part of the development rule which was previously accepted and saved in text file, if there is an error in this process user will be prompted to enter string input again and the same process would be followed till the entered input is not a part of the development rule.

In the system proposed in paper [3] they have suggested an online system for parsing sentences in Bangla language. They have broadly divided their work into three subparts the input part, the preprocessing part and finally parsing part. In input part they do accept a string data type from user in Bangla language which is further passed to preprocessing part in which the predefined module tokenizes the phrases properly and passes this module further to the parsing part in which they have already defined some standard algorithms which are applied on the tokenized part to extract grammatical meaning from those particular words from the dictionary. In the further steps they do rebuild the phrases in several ways and sent it to algorithm to verify the outputs reputedly.

The suggested system in paper [4] is to build a tree structure for CFG to make user friendly language interfaces for users so that they can relate to the platform more efficiently. They have discussed how the structure how the structure of CFG tree is a combination of tuple data structure of four elements namely NTPS where N and T stand for nonterminal and terminal symbols respectively, P stands for union of production regulations and S determines the starting symbol. They have used Hungarian as a test language for their proposed system which is comprised of very complex understanding of grammar. They have also used the database commands as a training set to their system to generate graphs based on it. This method proved to be an expert method to determine and study some of the internal structures of Hungarian grammar.

III. English Semantic Parsing

Parsing by Semantic is a method of tokenizing and assigning a proper structure to a sentence based on its meaning. The English grammar analytical rules and the interpretation technique are used to analyse the English semantic and syntactic organization of an user inputted phrase from the user standard input.

Our semantic parsing tool takes an English sentence as user input and generates an interpreted parse tree by implementing logical relatedness between the words. This parsing tool divides a token into individual components by applying a list of norms that characterise its organization and a token sequence to determine its correct grammar, which is specified by native grammatical rules. People with no particular language skills, who want to obtain textual relationships can readily understand and use representations because it provides straightforward

information about semantic correlations. Semantic relationships are used to indicate the relationships between the words and sentences. For the relevant, we presented our own parsing tool for extracting the tagged collections and context-free formatted grammatical representation.

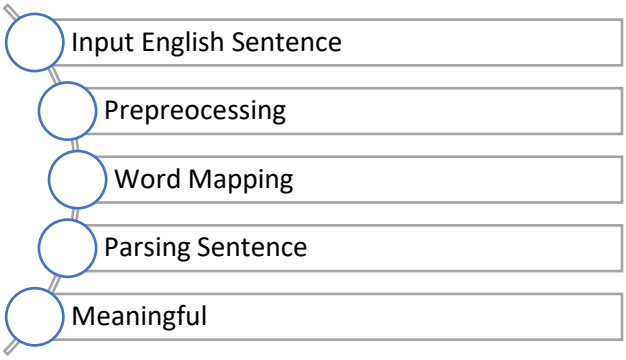

Fig. 1. Flowchart

A database stores pronoun from a sentence, singular-words, adverb , noun, plural from words , people, verbs, tense of the sentence, adjectives, and so on. Three phases make up the English Semantic Parser framework. The system tool takes an English sentence as input. The sentence is tokenized by the pre-processing program. Afterwards, using a language dictionaries, the interpretation algorithm is implemented to all tokens to identify the grammatical meaning of that word. Then, in various formats, it reassemble the phrase and pass it to the analyser. The analysed phrases are fed into the parser, which follows the CFG principles. The Phrase Modelling section will describe how to use a term dictionary and the suggested C-G-F principles to derive the semantics meaning of the phrase.

IV. Conclusion

We have used semantic representation and grammar to discover the English phrase among the tokens and created a parser as well. The primary intention was to develop a parser that could handle any form of English phrases. The purpose of this survey is to create a parser that can handle any sort of English sentence.

The following are the results of this survey:

a. For English Language Processing, we've created a parser module.

b. Different English sentences are entered into a lexicon.

c. To parse a statement in the English language, we used context-free grammar.

d. The parser module's output is expressed as a collection which can be displayed and handled with simplicity.

By altering the C-F-G regulation, we were able to construct a much more robust parser for the English language. The unambiguous representation of a phrase is the core issue of English language parsing. Furthermore, the researchers used the semantic rule to resolve word ambiguity. Our suggested parser was able to reach an accuracy of 91.57 percent. During the study, the authors observed that their efficiency was falling as they evaluated more uncertain words and sentences. Its reliability will improve in the future as more ambiguity problems are resolved.

References

[1] K. M. Azharul Hasan, Amit Mondal, Amit Saha Department of Computer Science and Engineering (CSE). “A Context Free Grammar and its Predictive Parser for Bangla Grammar Recognition” 13th International Conference on Computer and Information Technology (ICCIT 2010)

[2] VAYADANDE, KULDEEP. "Simulating Derivations of Context-Free Grammar." (2022).

[3] M. F. Mridha, Molla Rashied Hussein, Md. Musfiqur Rahaman, Jugal Krishna Das “A Proficient Autonomous Bangla Semantic Parser for Natural Language Processing”, ARPN Journal of Engineering and Applied Sciences, VOL. 10, NO. 15, AUGUST 2015,ISSN 1819-6608, pp 6398-6403.

[4] L.. Kovács, P. Barabás University of Miskolc/Department of Information Technology, Hungary “Experiences in building of context-free grammar tree”

[5] Vijay Kumar, Pankaj K. Sengar(2010), “Segmentation of Printed Text in Devanagari Script and Gurmukhi Script”, International Journal of Computer Applications Volume 3 – No.8, pp. 0975 –8887.

[6] Javed Ahmad Mahar, Ghulam Qadir MEMON(2010), “Rule Based Part of Speech Tagging of Sindhi Language", International conference on Signal Acquisition and processing.,pp.101-106.

[7] Pawan Goyal, Vipu Arora, Laxmidhar Behera (2009), “Analysis of Sanskrit text: Parsing and Semantic Relation”, Springer-Verlag Berlin Heidaelberg., pp. 200-218.

[8] Ms Vaishali M. Barkadeet. al. (2010), “English to Sanskrit Machine Translation Semantic Mapper”, International Journal of Engineering Science and Technology vol.2(10).

[9] Vayadande, Kuldeep, Ritesh Pokarne, Mahalaxmi Phaldesai, Tanushri Bhuruk, Tanmai Patil, and Prachi Kumar. "SIMULATION OF CONWAY’S GAME OF LIFE USING CELLULAR AUTOMATA." International Research Journal of Engineering and Technology (IRJET) 9, no. 01 (2022): 2395-0056.

[10] Vayadande, Kuldeep, Ram Mandhana, Kaustubh Paralkar, Dhananjay Pawal, Siddhant Deshpande, and Vishal Sonkusale. "Pattern Matching in File System." International Journal of Computer Applications 975: 8887.

[11] Vayadande, Kuldeep, Neha Bhavar, Sayee Chauhan, Sushrut Kulkarni, Abhijit Thorat, and

Yash Annapure. Spell Checker Model for String Comparison in Automata. No. 7375. EasyChair, 2022.

[12] Vayadande, Kuldeep, Neha Bhavar, Sayee Chauhan, Sushrut Kulkarni, Abhijit Thorat, and Yash Annapure. Spell Checker Model for String Comparison in Automata. No. 7375. EasaafyChair, 2022.

[13] Varad Ingale, Kuldeep Vayadande, Vivek Verma, Abhishek Yeole, Sahil Zawar, Zoya Jamadar. Lexical analyzer using DFA, International Journal of Advance Research, Ideas and Innovations in Technology, www.IJARIIT.com.

[14] Kuldeep Vayadande, Harshwardhan More,Omkar More, Shubham Mulay,Atahrv Pathak, Vishwam Talanikar, "Pac Man: Game Development using PDA and OOP", International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395-0056, p-ISSN: 2395-0072, Volume: 09 Issue: 01 | Jan 2022, www.irjet.net

[15] Kuldeep B. Vayadande, Parth Sheth, Arvind Shelke, Vaishnavi Patil, Srushti Shevate, Chinmayee Sawakare, "Simulation and Testing of Deterministic Finite Automata Machine," International Journal of Computer Sciences and Engineering, Vol.10, Issue.1, pp.13-17, 2022.

[16] Rohit Gurav, Sakshi Suryawanshi,Parth Narkhede,Sankalp Patil,Sejal Hukare,Kuldeep Vayadande," Universal Turing machine simulator", International Journal of Advance Research, Ideas and Innovations in Technology, ISSN: 2454-132X, (Volume 8, Issue 1 - V8I1-1268, https://www.ijariit.com/

[17] Kuldeep Vayadande, Krisha Patel, Nikita Punde, Shreyash Patil, Srushti Nikam, Sudhanshu Pathrabe, "Non-Deterministic Finite Automata to Deterministic Finite Automata Conversion by Subset Construction Method using Python," International Journal of Computer Sciences and Engineering, Vol.10, Issue.1, pp.1-5, 2022.

[18] Kuldeep Vayadande and Samruddhi Pate and Naman Agarwal and Dnyaneshwari Navale and Akhilesh Nawale and Piyush Parakh," Modulo Calculator Using Tkinter Library", EasyChair Preprint no. 7578, EasyChair, 2022

[19] Buczynski A., Wawer A.: Shallow parsing in sentiment analysis of product reviews. In: Proceedings of the Partial Parsing workshop at LREC, vol. 2008, pp. 14–18, 2