

A Parser: - Simple English Sentence Detector and Correction

Anup Mestry, Sangram Shende, Akshay Mahadik
 Department Of Computer Engineering, Kjsieit
 Ayurvihar Complex, Everard Nagar, Sion, Mumbai 400022
 Maharashtra, India
 Project Guide: - Prof. Shyamal Virnodkar

Abstract— Natural language processing is a study of mathematical and computational modeling of various aspects of language and the development of a wide range of systems. Natural language processing holds great promise for making computer interfaces that are easier to use for people, since people will be able to talk to the computer in their own language. For programming, however, the necessity of a formal programming language for communicating with a computer has always been taken for granted. We would like to challenge this assumption. We believe that modern natural language processing techniques can make possible the use of natural language to express programming ideas, thus drastically increasing the accessibility of programming to non-experts users.

In this paper we propose a way for English sentence detection and correction based on set of our production rules. It involves a function for detection and correction of simple English sentences. We take a simple sentence and evaluate it through production rules and gives appropriate result.

Keywords– Natural language processing, English language.

I. INTRODUCTION

Natural Language Processing is the computerized approach to analyzing text that is based on both a set of theories and a set of technologies. And, being a very active area of research and development, there is not a single agreed-upon definition that would satisfy everyone, but there are some aspects, which would be part of any knowledgeable person's definition.

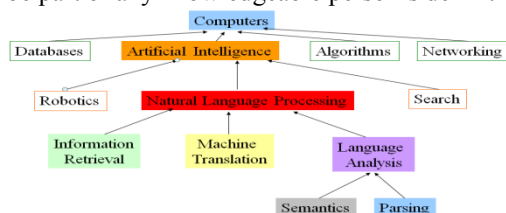


Figure 1: Classification of Computer Science

Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications.

NLP is a field of computer science and linguistics concerned with the interactions between computers and human (natural) languages. In theory, natural language processing is a very attractive method of human-computer interaction.

One of the most important modules of NLP is Parsing. This parser is used as inbuilt software in many application of NLP. In our project this parser will act as English sentence validator.

The English language offers a complex and ambiguous grammar that is readily understood by its natural users, but at times can be difficult to grasp by beginners/learners, and no less, by machines. NLP responsible for research and implementation of several techniques towards algorithmically analyzing and verifying the grammatical correctness of sentences in written English

II. BACKGRUOUND AND MOTIVATION

The traditional method of learning & teaching foreign language requires teacher to work 60 to 100 hours to bring an adult to a level at which she can function minimally the foreign language. By minimally functioning we just mean being able to use short, prepackaged phrases to communicate simple thoughts with long pauses and very foreign pronunciation. In general progress beyond this level requires a great deal more time and guidance in written form. Here we take the English language. Our project helps to understand foreign language material to verify grammatically infectious sentences this is possible because our algorithm will let the computer process the information much more conveniently. Although we normally think of language education as a task for schools and universities, which is certainty is, there is an enormous need for language training outside of the schools. It is convenient to consider three distinct sorts of language learners' academic, industrial, self-study as they differ in their interest for technical improvement and in their readiness to experiment.

Since natural language processing has become popular domain and many researchers are striving forward to design and implement their algorithms for natural language processing. Their implemented output it is not up to the mark here we decided to put forward our own logic and algorithm to process natural language sentences.

III. PROBLEM STATEMENT

Natural language processing and programming languages are both established areas in the field of computer science, each of them with a long research tradition. Although they are centered around a common theme – “languages” over the years. Based on the knowledge we are developing a parser which takes the simple English sentences as input. The input

is then taken through engine containing parsing rules which then processes the input and determines whether the sentence is correct grammatically or not. If entered sentence is grammatically wrong then parsing engine corrects it by using permutations and combination approach.

IV. REVIEW OF LITERATURE

Mohsin Ahmed and Milind Gandhedescribes the implementation of a natural language query processor based on the Pattern Matching technique [1]. The algorithm was to implement on IBM-PC/XT. A Study by Geller and Lesk about human factors in Menu versus Keyboard searches on a library database found that for a large database people preferred free form input. Here in this paper they have tried different approach to this problem by trying to provide to a natural language interface to the database system. They implement natural language query processor. Natural Language Query is written using PMP in prolog, and runs via prolog interpreter. The main aim behind using prolog is that recursive structure of the problem is easily represented in prolog. The fundamental data structures used are the sentence and the model. They have used standard dictionary as a part of natural language query.

However their algorithm was failing to understand the query written in natural language sometimes. For bypassing these semantic drawbacks they are providing echo questions and clarify the query correctly. Pattern matching paradigm is simplistic syntactic approach to natural language understanding and works well in limited applications. However natural language query cannot detect semantic contradictions as it does not use any data restriction information to detect semantic errors in the conversation.

Victor W. Zue described the concept which was developed based on the concept that spoken language offers an attractive alternative for human or computer interface, since speech is the most natural, efficient, flexible and economical means of communication among humans [2]. To provide this interface, however, speech recognition technology must be combined with natural language processing technology, so that the verbal input can not only be recognized, but also understood, and appropriate actions can be taken. On the input side, speech recognition must be combined with natural language processing in order to derive an understanding of spoken input. On the output side, some of the information that the user seeks as well as any clarification dialogue generated by the system must be converted to natural sentences and should be delivered as verbal responses. Since the implementation of this paper is quit successful but need some improvements in areas like working in real domains, spoken language generation, dialogue modeling, new word problem.

Fangju Wang introduces a technique for parsing incomplete queries written in natural language [3]. They say that formal query languages have limitations when applied to diversified users and large number of imprecise concepts. In earlier stages in their research, they incorporated natural language terms into formal query language structured query language, so that imprecise or vague search conditions can be used to retrieve geographic data. Then they developed a natural language user interface that could translate GIS queries in a natural language into formal query language

structured query language. The work was based on the possibility theory of fuzzy sets.

The core component of the technique is a fuzzy grammar that was aimed at handling missing grammatical constituents. They have developed a possibility distribution function which measures the possibilities for the variable to take specific values. This function helps us to find most possible values for the variable under the restriction. Their parsing algorithm is based on dynamic programming. Dynamic programming algorithm solves a problem by solving its sub-problems. The major advantage of using this technique is that computation cost is proportion to the number of rules which increases the performance of parser.

CONCLUSION

Natural language processing is a technique for parsing English language sentences. we make use of grammar for various for various purposes like sending letter to officials or any organization which requires data to be in formal format, we also use grammar when we are preparing for presentation, many people uses the grammar rules to make an email. English has not remained just for sake for speaking but it has reached to a level where people are not just required to make sentences, rather it weighs there importance by analyzing how people interact with each other in proper English. New production rules have been added in order to make parsing better. Grammar corrector and detector works for simple sentences helps to reduce for rereading of English sentences

Acknowledgment

Before presenting out our project entitled "A Parser: - Simple English Sentence Detector and Corrector", we would like to convey our sincere thanks to many people who guided us throughout the course for this project work. First, we would like to express our sincere thanks to our project guide Prof. Shyamal Virnodkar for providing precious help to carry out this paper. Finally, we would like to thank Mohsin Ahmed, Victor Zue, Fangju Wang for their research work in the field of natural language processing.

REFERENCES

- [1] Mohsin Ahmed and Milind Gandhe. Intelligent Natural Language Query Processor. Department of Computer Science and Engineering. Indian Institute of Technology, Bombay, INDIA 400076., in press.
- [2] Victor W. Zue. Human Computer Interactions Using Language Based Technology. Spoken Language Systems Group, Laboratory for Computer Science Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA, in press.
- [3] Fangju Wang. Parsing "Grammatically Incomplete" Natural Language Queries To Spatial Databases. Department of Computing and Information Science, University of Guelph, Ontario, Canada N1G 2W1fwang @snowwhite.cis.uoguelph.ca
- [4] Shi Guodong, GuYuwan, Sun Yuqiang, Wang Xiaokang, Yin Aling. Word-Lattice Parsing Parallel Algorithm. JiangSu University College of Electronic and Information Engineering, JiangSu, Zhenjiang212013, China.
- [5] GendLal Prajapati, On the inference of context free grammar based on bottom up parsing and search. Department of Computer Engineering Institute of Engineering & Technology, Devi Ahilya University Indore-452001, M.P., India
- [6] Earley, J. (1968). An Efficient Context-Free Parsing Algorithm. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA.
- [7] Harshad B. Prajapati, Vipul K. Dabhi, Logee Vaghela and Hiren Vataliya. Design and Implementation of XML Based Architecture for English to Hindi Language Sentence Translation.