Comparative Study of Rule Based Approach for Grammar Checker

Swapnali Deelip Baviskar
M.E Scholar
Department of Computer Engineering
SSBT's COET, Bambhori
baviskarswapnali5@gmail.com

Mr. Sushant S. Bahekar
Assistant Professor
Department of Computer Engineering
SSBT's COET, Bambhori
sushant.bahekar@gmail.com

Abstract

The grammar is defined as more than one words are connect together in text with a set of rules in natural language. Sentences must be syntactically as well as semantically correct in grammar checking. Rule-Based approach is employed in grammar checker as well as spelling prototype in Marathi language. Explanation, correction and detection of generic errors in sentence structure as well as punctuation are done by rule matcher module. For identifying sentences with wrong structures, phrase chunks, tokens or POS tags are taken into consideration for a language model. Grammar checker makes use of it. The phrase chunk model gives best results by normalizing chunk length in grammar checker. An overall accuracy is depicted in the system analysis. This prototype has several application in the world, the best being a web application.

Keywords: *Natural Language Processing, Grammar Checker, POS Tags.*

1. Introduction

A natural language is a language used by humans for writing and speaking purpose. Natural language processing (NLP) is a field in computer science, artificial intelligence and the expressive which mainly focuses on the communications between human languages (natural language) and the computer. Natural languages are used for communication purpose. Today is the era of information technology. Everyone is using web to share and find information. But, the information is present in natural languages. Human make mistakes while writing, which can be divided into two kinds: spelling mistakes and grammatical mistakes. Spelling mistakes are mistakes where the given word does not belong to vocabulary of the language whereas grammatical mistakes correspond to the mistakes where the sentence does not confirm to rules forced on the languages grammar. A Grammar checker is a software or program or idea which is designed to spot grammatical mistakes in the given text with the ability of either automatically correcting the mistakes or suggesting correct options.

For grammar checking in language there are widely using three methods: Syntax-based checking, statistics-based checking and rule-based checking. Each sentence is fully parsed by using syntax based grammar checking method. When the syntactic parsing fails then text is examine incorrect. In statistics-based approach, an annotated corpus and the probability and the frequency are built for POS tag sequences. When the POS-tagged text contains POS series with frequencies less than some method then there is the text is considered incorrect. The labeled training corpus is necessarily learns the rules in The statistics based approach. Except that the rules must be made up manually in rule-based approach. Rule-based approach is similar as a statistics-based approach.

Rule-Based is a method of representing huge documents in a condensed form without affecting the meaning of the text. So, it is a method of presenting the most important content of a document to the reader from a set of unstructured document/documents. The main motivation comes from the fact that much work has been done for grammar checker for different Indian languages such as Hindi, English, Urdu and Punjabi as well as foreign language but no work has been done in Marathi language as its grammar is quite difficult to implement. Here, an approach to check grammar for Marathi sentences is introduced.

The objectives of this project are:

- To resolve errors in punctuation & sentence structure with an overall accuracy.
- To improve the efficiency and accuracy of Marathi grammar.
- To improve the performance of grammar checker for Marathi sentences in terms of precision, recall and feature score.

2. Literature Survey

- Asanilta Fahda et. al., in [1] presented the rule-based approach for a combination of rules and statistical methods are made prototype for an Indonesian spelling and grammar checker. The rule matcher module works on 38 rules which detect, correct, and find common errors in punctuation, word choice, and spelling. To recognize sentences with incorrect structures they are uses the trigram language model for grammar checker from POS tags, phrase chunk or tokens. 83.18% is the complete accuracy of the system is based on document analysis.

- Shashi Pal Singh et. al., in [2] focus on one should aim to minimize the errors while using the language. Lesser is the number of mistakes, better will be the communication. To help in achieving this goal, they are creating a frequency based spell checker and a rule based grammar checker for English language. The grammar checker focuses on identifying and correcting tense related mistakes.

- Nivedita S. Bhirud et.al., in [3] described that review look at past, present and the future in the present context for development of various Natural Language grammar on the till date. The grammar checkers of a few Indian languages as well as foreign Languages discussed with the different characteristics are conclude in this survey. They are covers the grammar checkers for various languages and approaches, methodologies and performance evaluation and also common grammatical errors which would be introduced to new tool and system as a whole along with the key concept and grammar checker internals. They observed that professionally grammar checker is available for English language, while for most other languages, the work is in progress. The grammar checker for Marathi language could not have been reproduced.

- Lata Bopche and Gauri Dhopavakar in [4] describes a method for Hindi grammar checker. This system made up to rule based system and morphological analysis for lexicon. The basic process like POS tagging, tokenization and morphological analysis by passed to the input text. The result is only for simple sentence in there system. The input sentence have the same number of words only checks by the system.

- D. Naber in [5] described a possible errors in text and return in list. Each sentence is split into chunks, e.g. noun phrases is allocated by its part-of-speech tag and detect errors. The text is matched by the checker's predefined error rules. The rule matches contains an errors at the situation of the match. Errors as patterns of words, part-of-speech tags and chunks explain by rules. A description of the error is described by the each rule, which is shown to the user. To allow the users to write their own rule, still it is sufficient to identify many errors by rule-based approach system.

- Swati Ramteke et. al., in [6] developed lexicon Parser for Devanagari script (Hindi), how the parse tree is create by using the semantic representation and using the grammar show the relationship between tokens and sentence is parsed into tokens for in Hindi language. To explain the disambiguated of words they used Rule based approach. The algorithms are used to develop and implemented for Devanagari text are: tokenization and tagging algorithm. Lexicon parser is provide the accuracy of 89.33%. The system is observed by experiment that when they tested more ambiguity sentences and sentences of future tense then their correctness was low. On testing, some tenses such as simple past and present tenses has very high accuracy.

- Mandeep Singh et.al., in [7] implemented the system executes morphological analysis using the rule based system for lexicon and POS tagging and phrase chunk. The suggestion provided for detecting the grammatical errors in Punjabi texts used the grammar checking software. This system's prime attraction is that the detected errors are presented in detailed form. It comes with the suggestion as well. The compound sentences as well as complex sentences are dealt with. Specially designed error detection rules are supported by system for Punjabi texts. Agreements and order of word in phrases generate several grammatical errors.

- Jahangir Md et. al., in [8] has done research on correctness of a sentence. Analysis of words based on n gram and POS tags are taken into consideration to check correctness of a sentence. With the help of POS tagger, every single word of sentence is allotted tag by the system. Tag sequence's probability is determined with the help of gram analysis. For sequence to be correct, the probability must be one or more. Bangla as well as English languages are compatible with the system. POS tagging is very vital as system is dependent on it. Manually tagged and automatically tagged sentences are checked by the author. After execution, the results looked very promising for languages like Bangla when compared with English. This is due to brown corpus method as it consists huge number of compound sentences.

- Research study has been done on two pass parsing approach by H. Kabir et. al. [9]. It is mainly employed in analyzing input text. Redundancy is one of the critical obstacles in the phrase structure. Above approach tries to reduce that redundancy effectively. Redundancy in grammar rules which are utilized in sentence analysis is also rectified. When a case result in failure, reparsing of tree is done using movement rules. POS guesser and Morphological disambiguation's module may result in system not working properly. Rest of the time, system runs properly. Structural mistakes and grammatical mistakes are checked by the grammar checker. Mostly these mistakes occur in declarative sentences. The grammar checker also try to suggest error corrections.

Table 2.1 shows the literature survey of Grammar Checker.

Paper	Method	Result	Future Scope
Asanilta Fahda et. al., in [1]	Rule Based	It was able to check errors in punctuation, word choice, spelling, and sentence structure	The statistics-based components can be further improved by creating corpuses of

		with an overall accuracy of 83.18% by using rules with co - occurrence ranking, and a phrase chunk normalized trigram model grammar checker.	Indonesian spelling and grammar errors.
Shashi Pal Singh et.al., in [2]	Rule Based	The grammar checker works on the rule-based approach whereas the spell checker is based on frequency based approach.	The accuracy of the grammar can be further increased by refining the grammar based rules. The application of grammar checking can be applied to the various aspects of the grammar apart from the tenses in English language, such as the Active and Passive voice, Direct and Indirect speech, etc.
Nivedita S. Bhirud et.al., in [3]		The aim of the survey was to study various Grammar Checkers on the scale of their features such as types of grammar errors, weaknesses and evaluation.	The future scope for developing grammar checkers for uncovered languages with feasible approach. And there future research work aims to develop grammar checker for Marathi Language.
Lata Bopche and Gauri Dhopavakar in [4]	Rule Based	The system is designed to work on the literary Style Punjabi text with SOV (Subject Object Verb) sentence structure. It can detect any agreement errors in compound or complex sentences also.	
Swati Ramteke et. al., in [6]	Rule Based	Rule based approach used to resolves the disambiguity of words. The accuracy of 9.33% was achieved from Lexicon parser.	The accuracy increase when they tested more ambiguity sentences and sentences of future tense.

3. Conclusion

In this survey, several grammar checking methods have been reviewed. Grammar checker internals, different approaches along with important concepts also have covered in the review. The main goal of survey was about studying different features for different grammar checkers. Features includes parameters like weakness, types of grammar errors, weakness etc. On that scale, we could find out about the effectiveness and efficiency of grammar checker.

4. Future Scope

In this paper, we done comparative study and analysis of various grammar checker techniques but techniques have some pros and cons, there is necessity to develop a such technique, that overcome all these disadvantages with a proper grammar checker and increase accuracy in Marathi sentences.

Acknowledgement:

First and foremost, I would especially like to thank my adviser and guide Mr. S. S. Bahekar, for being a wonderful instructor and always being in bright spirits. I am incredibly grateful to Principal K. S. Wani, Head of the department Prof. Dr. Girish K. Patnaik and the Professors in the M.E. Computer Engineering Program at SSBT's COET,Jalgaon. Without their time and support, I would not have been able to complete this work. Last, but not least, I like to thank my family and friends who also supported me throughout my studies.

References

[1] Asanilta Fahda, Ayu Purwarianti, "A Statistical and Rule-Based Spelling and Grammar Checker for Indonesian Text", 2017 International Conference on Data and Software Engineering, 2017.

[2] Shashi Pal Singh, Ajai Kumar and Lenali Singh, "Frequency based Spell checking and Rule based Grammar Checking", in proceedings of the 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) - 2016.

[3] Nivedita S. Bhirud, R.P. Bhavsar and B.V. Pawar, "Grammar Checkers for Natural Languages: A Review", International Journal on Natural Language Computing (IJNLC) Vol. 6, No.4, August 2017.

[4] Lata Bopche and Gauri Dhopavakar, "Rule Based Grammar Checking System for Hindi", Journal of Information Systems and Communication ISSN: 0976-8742 and E-ISSN: 0976-8750, Volume 3, Issue 1, 2012.

[5] D. Naber, "A rule-based style and grammar checker", Diplomarbeit Technische Fakultat, Universitat Bielefeld, Germany 2003.

[6] S.Ramteke, K.Ramteke, Rajesh Dongare ,"Lexicon Parser for syntactic and semantic analysis of Devanagari sentence using Hindi wordnet", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 4, April 2014.

[7] Mandeep Singh, Gurpreet, Sharma and Shiv, "A Punjabi Grammar Checker", Punjabi University. 2nd international conference of computational linguistics: Demonstration pa-per. 2008. pp. 149 132.

[8] Jahangir Md, Uzzaman, Naushad, Khan, Mumit, "N-Gram Based Statistical Grammar Checker For Bangla And English", Center for Research On Bangla Language Processing. Bangladesh, 2006.

[9] H. Kabir, S. Nayyer, J. Zaman, and S. Hussain, "Two Pass Parsing Implementation for an Urdu Grammar Checker."