

# A Review Paper on Text Corrector Using Finite Automata

Mrugakshi Deshmukh, Shreyash Deshmukh, Devansh Jangid, Dhanashree Dewalkar and Rushikesh Dighole

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

March 28, 2022

# A Review Paper on Text Corrector using Finite Automata

Mrugakshi Deshmukh Artificial Intelligence and Data Science Vishwakarma Institute of Technology, Pune,411037, Maharashtra. mrugakshi.deshmukh20@vit.edu Shreyash Deshmukh Artificial Intelligence and Data Science Vishwakarma Institute of Technology, Pune,411037, Maharashtra. shreyash.deshmukh20@vit.edu Devansh Jangid Artificial Intelligence and Data Science Vishwakarma Institute of Technology, Pune,411037, Maharashtra. devansh.jangid20@vit.edu

Dhanashree Dewalkar Artificial Intelligence and Data Science Vishwakarma Institute of Technology, Pune,411037, Maharashtra. dhanashree.dewalkar20@vit.edu

Abstract—Time plays an important role in everyone's life, and for some, it is not easy to spare it. In scenarios where a person has to meticulously keep track of spelling legibility and websites that are to be listed, time runs short. And it is therefore important to rely upon tools like spelling checkers ensure minimum human error. Use to of Deterministic Finite Automata is undertaken, to identify which language a given string of characters is from, and also to verify whether the order of strings represents the structure of a valid URL. When a DFA is being run, it will come to the end of the string after a while. And then will result in a match, if possible.

Keywords — Deterministic Finite Automata, Levenshtein edit-distance, Spell checker, NonDeterministic Finite Automata

#### I. INTRODUCTION

The most important part of human life is language, which can be represented in either spoken or textual content. Language has a significant role in Rushikesh Dighole Artificial Intelligence and Data Science Vishwakarma Institute of Technology, Pune,411037, Maharashtra. rushikesh.dighole20@vit.edu

document writing. The misspelling makes the content more difficult to read and reduces the text's informational value. While composing the paper, some mistakes may occur due to human error, such as a slip of the finger on some other key or the adjacent key on the keyboard. Many web apps have a spell checker, which is a highly valuable tool for users. It may offer options in a dropdown menu so that the tool will type the rest of the word that the user intends to type. This tool assists the user in correcting errors in document text writing. Most of the spell checker systems which have been designed used Deterministic Finite Automata (DFA)[3][16][18]. When discussing the spell checker, the two primary keywords that come to mind are error detection and error repair, which implies that the first task is to discover the erroneous words in the document and then provide appropriate suggestions for the wrong word. The basic premise behind a spell checker is that each word in the document is looked up in the dictionary, and if it isn't found, it is assumed to be an incorrect term. То limit the amount of misspelled words, the system searches the

dictionary for words that are most similar to the error word and displays them in a list for the user to choose from.

#### **II. LITERATURE REVIEW**

Vibhakti Bhaire et al.[1] proposed an application that detects and corrects users' writing errors. It scans the word and compares it to the dictionary which are correctly spelled). (words To accommodate morphology. an additional language-dependent procedure (Edit distance) is introduced. It also has autocorrect and multiword suggestion features. Java was used as a programming language, and the Net-Bean IDE was used as an IDE.

Recently, Anita Patil et al.[2] has conducted a study on the various strategies used by spell checkers for Indian languages. It divides spelling errors into two categories: typographic errors (non-word errors) and cognitive errors (Real Word errors). Ngram and Dictionary look-up were among the error detection methods described in the paper. Finally, it discussed ways to correct errors such as Edit distance, similarity keys, rule-based procedures, and N-gram techniques. They found out that many spell checkers did not include Marathi language.

In Joseph Marvin R. Imperial and team's implementation [3], DFA was used along with Levenshtein edit-distance method, to parse strings and report the desired result. A separate wordlist was prepared for training, containing different words and their various verb-forms. The wordlist was then parsed using RegEx to convert it to an understandable (by the DFA) expression. After converting the RegEx expression to NFA, the Levenshtein Edit-Distance Feature was used to suggest possible spellings.

Viny Christanti Marwadi and company [4] implemented the DFA in such a way that the strings were first subjected to pre-processing, which included the steps case folding (reduce every single alphabet to lowercase, here), removing unneeded characters like numeric characters and punctuation symbols, and finally separation of remaining characters into words with the help of spaces as indicators. After preprocessing the user input string is verified with the help of a prepared dictionary, where the given string is looked up. Calculations using Levenshtein edit-distance feature are then performed to calculate exactly how many differences there are between any two strings. Finally, the user is taken to a suggestion's page where the errors are highlighted and the user can take or reject the suggestion.

Tommi A Pirinen's et al. [5] paper presents methods for spelling evaluation using Lexical Automation. The main aim is to simply recommend different algorithms that can be used to build and create spell checkers for the end user's convenience. Finite State methodology has been suggested using a finite state accelerator using regular weights to symbolise spelling errors and map them to their corresponding correct string(s). The data sets being used are from Wikipedia. This data was used to give an example of how to train the model which contains information on the frequency of different words (taken from Wikipedia).

Amanjot Kaur et al. [6] has an aim to make a spelling checker for Punjabi language. In her paper he tries and explain how different techniques like Dictionary lookup technique or Edit Distance Technique to make spell checker for Punjabi Language

The paper [7], Language Independent Text Correction using Finite State Automata by Ahmed Hassan, Sara Noeman and Hany Hassan attempts to propose an innovative way for automatic correction of spelling mistakes by using finite state automata to provide users corrections within a certain edit limit from the original word. This way of doing spell correction only needs a dictionary and text data to build the model. The system proposed by the authors consists of 3 main phases of the system, the first one is detection of misspelled words. Second is generating candidate corrections for them and the last step is ranking those corrections.

Prabhakar Gupta et al. [8] an employee at Amazon, proposed a real-time spell-checking system that not only checks spellings but also suggests context sensitive corrections. He also claims that his system can be easily extended to support other languages with minimal language related changes

This paper [9], Aadil Ahmad Lawaye et al., mainly focuses on the three parameters i.e. Creation of a dictionary using Lexicon, Error Detection, Error Correction and Replacement with the goal of detecting non-real word errors and the system successfully detected 80% of the errors and provided 85% of correct suggestions.

The paper [10], Applications of Finite Automata Representing Large Vocabularies by Authors Claudio L. Lucchesi and Tomasz Kowaltowski attempt to propose the concept of compressing a big vocabulary of words into a finite automaton that is minimally acyclic and deterministic. The corporation possessed a very complete machine readable vocabulary of roughly 206,000 words, but it was having major difficulties finding a compact representation.

Victoria J. Hodge, with her colleagues, [11] proposed a spelling checker using the Novel Binary Neural Approach. Their system is only aimed toward isolated word error correction. They assign priority numbers to the suggested words according to the similarity with the user's entered word.

### IV. ADVANTAGES AND DISADVANTAGES

## 1. Advantages

- 1. Simple to recognize.
- 2. Detects the majority of spelling and typing problems.
- 3. Suggests that a space between two consecutive words was omitted during typing by indicating that the word is misspelt.
- 4. Provides a list of possible spellings, one of which may be correct.
- 5. Accuracy is improved, and time is saved.

### 2. Disadvantages

- Does not distinguish homonyms (for example, by – buy, their -there – there, too – to – two) and so does not detect improper homonym spelling.
- 2. Proper nouns may be mistakenly identified as spelling errors (proper names of persons or places are not in the Spell-Checker dictionary, as they are not usually found in a traditional Dictionary). Students should pick "Skip" or "Ignore" on the dialogue box in this case.
- 3. Detects errors when words are spelled in a language other than the program's default—for example, Canadian spelling (e.g. color) in an American default (e.g., color). In this scenario, see if a Canadian dictionary may be set as the default vocabulary in the word processing program.

#### V.CONCLUSION

This survey analyzed different spell-checking systems using a variety of key concepts and approaches. Studies of these systems are done on the scale of their features like language and spelling errors. Observations say that most of the approaches and systems are for the English language specifically. Hence the new systems to be proposed should build systems which are adaptive to different languages and error types.

#### REFERENCES

- Vibhakti V. Bhaire, Ashiki A. Jadhav, Pradnya A. Pashte, Mr. Magdum P.G, "Spell Checker", 2015, Vol 5
- [2] Anita A. Patil, Priti R. Sharma," Study and Review of Selective Spell-Checking Approaches",2019, Vol8
- [3] Joseph Marvin R. Imperial, Czeritonnie Gail V. Ya-On, Jennifer C. Ureta, "An experimental Tagalog Finite State Automata spellchecker with Levenshtein edit-distance feature", 2019
- [4] Viny Christanti Marwadi, Niko Susanto, and Dali Santun Naga, "Spelling Correction for Text Documents in Bahasa Indonesia Using Finite State Automata and Levenshtein Distance Method Spelling Correction for Text Documents in Bahasa Indonesia", 2018
- [5] Tommi A. Pirinen, Krister Lindén, "Finite State Spell Checking with Weighted Language and Error Models-Building and Evaluating Spell-Checkers with Wikipedia as Corpus".
- [6] Amanjot Kaur, Dr. Paramjeet Singh, Dr. Shaveta Rani, "Spellchecking and Error Correcting System for text paragraphs written in Punjabi Language using Hybrid approach", Nov 2015, Vol. 2, International Journal of Advanced Research in Science, Engineering and Technology
- [7] Ahmed Hassan, Sara Noeman, Hany Hassan, "Language Independent Text Correction using Finite State Automata" IBM Cairo Technology Development Center.
- [8] Prabhakar Gupta, "A context sensitive real time Spell Checker with language adaptability", Oct 2019.
- [9] Aadil Ahmad Lawaye, Prof. Bipul Syam Purkayastha, "Design and Implementation of Spell Checker for Kashmiri", July 2016, Vol. 5, International Journal of Scientific Research.
- [10]Claudio L. Lucchesi, Tomasz Kowaltowski, "Applications of Finite Automata Representing Large Vocabularies", 1992 Brazilian National Council for Scientific and Technological Development

- [11] Victoria J. Hodge and Jim Austin, "A Comparison of Standard Spell-Checking Algorithms and a Novel Binary Neural Approach", 2003, Vol. 15
- [12]Daniel Hládek, Ján Staš, and Matúš Pleva," Survey of Automatic Spelling Correction",2020
- [13]Tommi A Pirinen, Sam Hardwick, "Effect of Language and Error Models on Efficiency of Finite-State Spell-Checking and Correction", 2012 10th International Workshop on Finite State Methods and Natural Language Processing.
- [14]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.
- [15]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.
- [16]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.
- [17]Vayadande, Kuldeep. "Simulating Derivations of Context-Free Grammar." (2022).
- [18]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.
- [19]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

- [20]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.
- [21]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/
- [22]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.
- [23]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