



Design and Implementation of Online Punjabi Spell Checker Based on Dynamic Programming

Ritika Mishra*

M.Tech Research Scholar

Department of Computer Science Engineering
SGGSWU, Fatehgarh Sahib, India.

Navjot Kaur

Assistant Professor

Department of Computer Science Engineering
SGGSWU, Fatehgarh Sahib, India.

Abstract— *Spell checker is a supercomputer application that analyzes possible misspellings in a text by referring to the accepted spellings in a database. Enormous amount of work has been completed in the part of English language. Punjabi is the official language of Punjab state in India. In Punjabi language, there is very small amount of work is completed in this region. There are only two spell checkers for Punjabi language i.e. “Akhar” and “Sudhaar”. Akhar is paid software that is not available free for its use to everybody and Sudhaar spell checker is a desktop application. This paper describes the development and working of online Raftaar Punjabi spell checker and also developed a proposed algorithm for the correction of wrong words, This System gives the result accuracy as 80% according to the research work for Punjabi words. It gives nearby result up to 80% of words tested in this thesis. It gives results for rest of 20% but not the best possible correct word was displayed on the top of the correct word list from the database.*

Keywords— *Spell Checker, Dynamic Programming, Real-word errors and Non-word errors.*

I. INTRODUCTION

Spell checker is a supercomputer application that analysis possible misspelling in a text by referring to the accepted spellings in a database. In other words, Spell Checker is an electronic lexicon in a word processor that can be used to hold misspelled words. Spell Checker has three mechanism: An error detector that detects misspelled words, a candidate spelling initiator that gives spelling suggestions for the detected misspelled words and an error corrector that choose the finest correct spellings out of the list of candidate spelling. Spelling errors can be partitioned into different categories, that is, real-word error and non-word error. A real-word error those error words that are acceptable word in the lexicon and non-word error those error words that cannot be found in the lexicon [1].

Dynamic Programming is a method for solving complex problems by breaking them down into simpler sub-problems. In general, to solve a given problem, we need to solve different parts of problem (sub-problems), and then combine the solutions of the sub-problems to reach an overall solution [2]. Often, many these sub-problems are really the same. The dynamic approach seeks to solve each sub-problem only once, thus reducing the number of computations. This is especially useful when the number of repeating sub-problems is exponentially large. Top-down dynamic programming simply means storing the results of certain calculations, which are later used again since the completed calculation is a sub-problem of a larger calculation. Bottom-up dynamic programming involves formulating a complex calculation as a recursive series of simpler calculations.

II. PROPOSED WORK

Raftaar is a recently developed spell checker for Punjabi language. It is an online application developed in an ASP.NET language with Sql Server 2005. The main features of Raftaar Punjabi spell checker are large database, online application, easy to operate, email and printing option.

- **Home Page**

We develop a new web page for introducing the newly developed Raftaar Punjabi Spell Checker as shown in the figure 1.1.



Figure 1.1: Snapshot of Home Page

- About Us

This page describes the overview of spell checker as shown in the figure 1.2.



Figure 1.2: Snapshot of About Us Page

- Working

This page describes the working of our spell checker as shown in the figure 1.3.

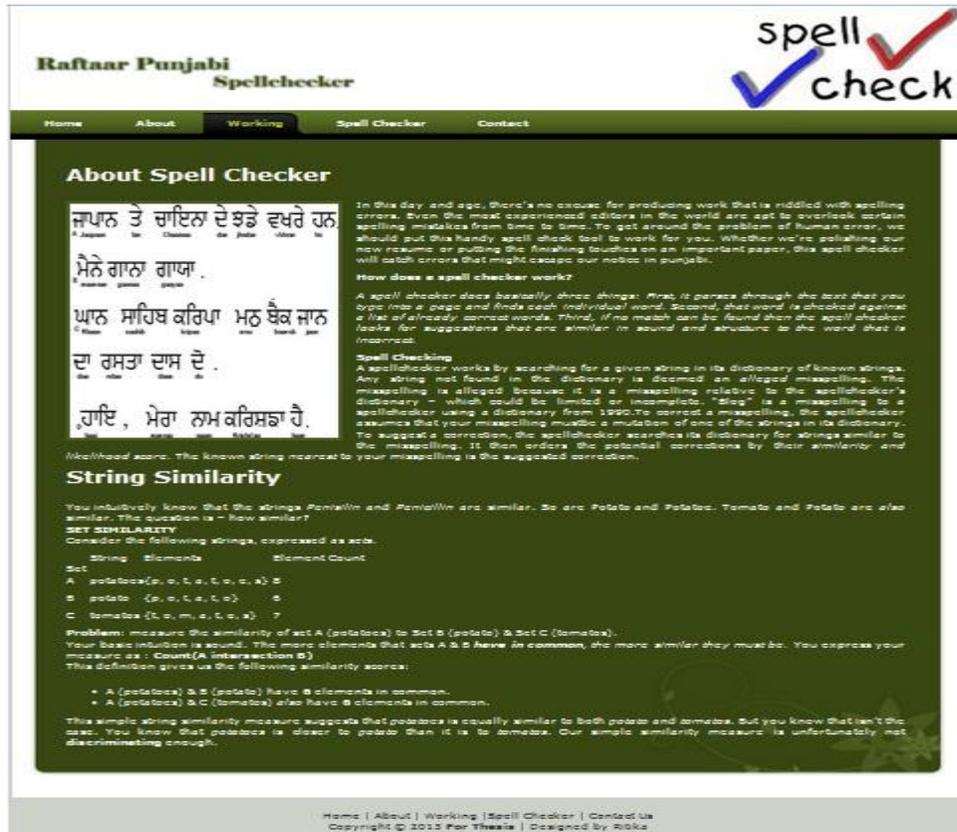


Figure 1.3: Snapshot of Working of Spell Checker

- Spell Checker

When user clicks on Spell Checker, the window appears which contains a text area as shown in the figure 1.4.. The user can browse the text file via browse button or enter the text manually and user can check the spelling errors via spell check button.



Figure 1.4: Snapshot of Spell Checker

- **Contact Us**

User can submit their feedback via submit button as shown in the figure 1.5. User can also give any suggestion through email id.

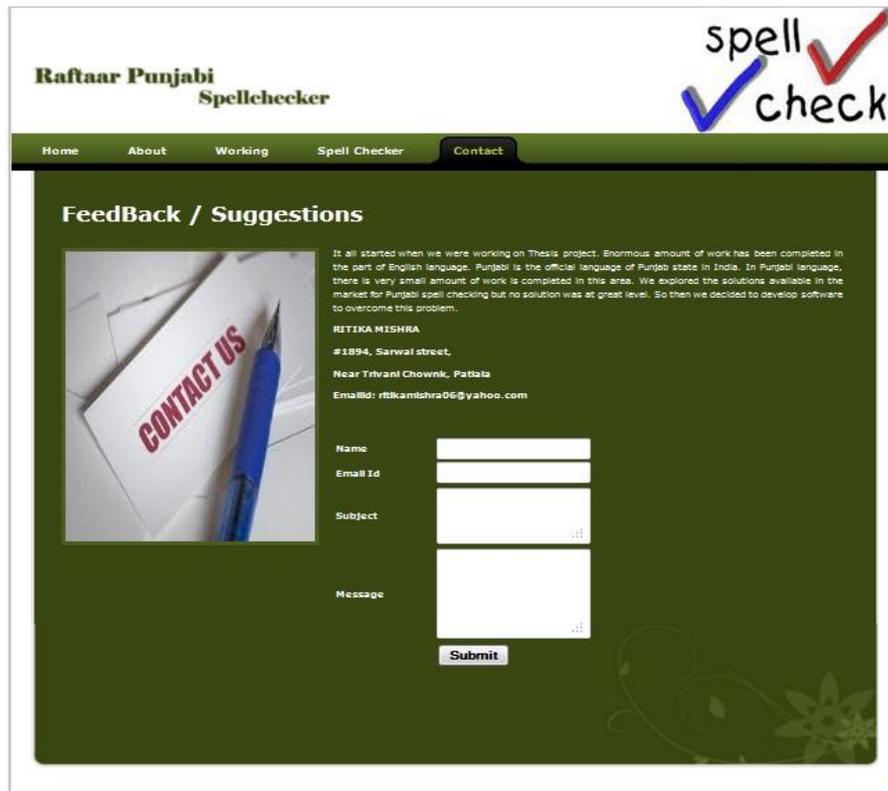


Figure 1.5: Snapshot of Contact Us Page

Algorithm

- Step 1: Parse the paragraph into tokens.
- Step 2: Store all the tokens in a string array paragraph.
- Step 3: Select the first word of the paragraph as $i=1$.
- Step 4: Repeat steps 5 to 7 until last word of paragraph traverse as $i \leq n$, where n is the number of words in a paragraph.
- Step 5: Initialize the flag status as TRUE.
- Step 6: Compare the selected word of the paragraph with each word of the database.
- Step 7: If the comparison result is TRUE.
 - (a) Change the flag status as FALSE.
- End of Step 4 Loop
- Step 8: If the flag status is TRUE.
 - (a) Compare the each letter of the paragraph word with one by one letter of all the words of the database.
- Step 9: If the comparison result is TRUE.
- Step 10: Increment the counter by 1.
- End of Step 9 if statement.
- End of Step 8 if statement.
- Step 11: If the difference between numbers of letters of a selected word of a paragraph and counter is equal to 1. Step 12: Insert the wrong word and write word in the database.
- End of Step 11 if statement.
- Step 13: View all the options of correct word with the wrong word from the database.

III. RESULTS AND DISCUSSIONS

We perform testing on different examples corresponding to our algorithm.

Test Case: 1

In test case 1, gau is an input data i.e. incorrect word as shown in the figure 1.6 (a). After clicked the spell check button, dictionary gives the most nearest matching correct words. The user can select the own relevant correct word. Then input data is replaced with the correct word as shown in the figure 1.6 (b).

ਲੜੀ ਨੰ.	ਗਲਤ ਸ਼ਬਦ	ਸਹੀ ਸ਼ਬਦ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ
1	ਗਉ	ਗਉ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ
2	ਗਉ	ਗਉਸ਼ਾਲਾ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ

Figure 1.6 (a) Incorrect Word

ਲੜੀ ਨੰ.	ਗਲਤ ਸ਼ਬਦ	ਸਹੀ ਸ਼ਬਦ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ
1	ਗਉ	ਗਉ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ
2	ਗਉ	ਗਉਸ਼ਾਲਾ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ

Figure 1.6 (b): Correct Word

Test Case: 2

In test case 2, we can browse a text file. The text will be shown in a text area after click on the view text button as shown in the figure 1.7 (a). After clicked the spell check button, dictionary gives the most nearest matching correct words. Then input data is replaced with the correct word as shown in the figure 1.7 (b).

Upload .txt File No file selected. upload/a1.txt File uploaded successfully!

ਮੈਂ ਸਕੂਲ ਗਇ

[Print](#) [E-Mail](#)

[Home](#)

ਲੜੀ ਨੰ.	ਗਲਤ ਸ਼ਬਦ	ਸਹੀ ਸ਼ਬਦ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ
1	ਮੈਂ	ਮੈਂ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ
2	ਸਕੂਲ	ਸਕੂਲ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ
3	ਗਇ	ਗਈ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ
4	ਗਇ	ਗਏ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ

Figure 1.7 (a): Incorrect Word

Upload .txt File No file selected. upload/a1.txt File uploaded successfully!

ਮੈਂ ਸਕੂਲ ਗਈ

[Print](#) [E-Mail](#)

[Home](#)

ਲੜੀ ਨੰ.	ਗਲਤ ਸ਼ਬਦ	ਸਹੀ ਸ਼ਬਦ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ
1	ਮੈਂ	ਮੈਂ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ
2	ਸਕੂਲ	ਸਕੂਲ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ
3	ਗਇ	ਗਈ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ
4	ਗਇ	ਗਏ	ਬਦਲਣਾ	ਸ਼ਬਦ ਕੋਸ਼ ਵਿੱਚ ਜਮਾ

Figure 1.7 (b): Correct Word

IV. Conclusion and Future Scope

In our Research work, we have developed an online Raftaar Punjabi spell checker and also developed a new proposed algorithm for the correction of wrong words according to the dictionary. The main features of Raftaar Punjabi spell checker are large database, online application, easy to operate, email and printing options. This System gives the result accuracy as 80% according to the research work for Punjabi words. It gives nearby result up to 80% of words tested in this thesis. It gives results for rest of 20% but not the best possible correct word was displayed on the top of the correct word list from the database. In this Research work, the word is not given the highlighter for wrong words. The future scope for this project as the words highlighted with red highlighter which are not correct according to the dictionary. For further research, some grammatical rules like the combinations of noun, verb, and adverb may be added.

Acknowledgment

I am thankful to my guide Navjot Kaur, Assistant Professor in Department of Computer Science Engineering for their valuable support. Also I am thankful to my Department for the technical support.

References

- [1] Rupinderdeep Kaur and Parteek Bhatia, “*Design and Implementation of SUDHAAR-Punjabi Spell Checker*,” International Journal of Information and Telecommunication Technology, Vol. 1, Issue 15 May, 2010.
- [2] S. Dasgupta, C.H. Papadimitriou, and U.V. Vazirani, 'Algorithms', p173, available at <http://www.cs.berkeley.edu/~vazirani/algorithms.html>.
- [3] Neha Gupta &PratisthaMathur,“*Spell Checking Techniques in NLP: A Survey*,” International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, Issue 12, December 2012.
- [4] Gurpreet Singh Lehal, “*Design and Implementation of Punjabi Spell Checker*”, International Journal of Systemics, Cybematics and Infomatics, 2007.
- [5] G S Lehal&MeenuBhagat, “*Spelling Error Pattern Analysis of Punjabi Typed Text*”, In Proceedings of International Symposium on Machine Translation, NLP and TSS, pp. 128-141, 2007.
- [6] Jesus Vilares& Manuel Vilares, “*Managing Misspelled Queries in IR Application*,” Issue 8, October 2010.
- [7] Youssef Bassil& Mohammad Alwani, “*Context-sensitive Spelling Correction using Google Web IT 5-Gram Information*,” Department of Computer and Information Science, Vol. 5,No.3, May 2012.
- [8] Amit Sharma &Pulkit Jain, “*Hindi Spell Checker*”, Indian Institute of Technology Kanpur, April 17, 2013.
- [9] MeenuBhagat, (2007), “*Spelling Error Pattern Analysis of Punjabi Typed Text*”, Thesis Report, Thapar University, Patiala.
- [10] F.J. Damerau (1964), “*A Technique for Error Detection and Correction of Spelling Errors*”, Communication ACM, pp. 171-176.
- [11] Monisha Das, S. Borgohain, JuliGogoi, S. B. Nair (2002), “*Design and Implementation of a Spell Checker for Assamese*”,lec, pp. 156, Language Engineering Conference (LEC'02).
- [12] Morris, Robert & Cherry, Lorinda L, “*Computer Detection of typographic errors*”, IEEE Trans Professional Communications, vol. PC-18, no. 1, pp 54-64, March 1975.
- [13] R.E. Gorin (1971), “*SPELL: A spelling checking and correction program*”, Online documentation for the DEC-10 computer.
- [14] K. Kukich (1992) “*Techniques for automatically correcting words in text*”. ACM Computing Surveys. 24(4): 377-439.
- [15] Peterson James (1980), “*Computer Programs for Detecting and Correcting Spelling Errors*”, Computing Practices, Communications of the ACM.