

**INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH
TECHNOLOGY****SPELL CHECKING AND ERROR CORRECTING SYSTEM FOR TEXT
PARAGRAPHS WRITTEN IN PUNJABI OR HINDI LANGUAGE USING HYBRID
APPROACH****Amandeep Singh*, Prof. Meenakshi Sharma**

M-tech student, CSE Department, SSCET Badhani Pathankot

HOD ,CSE Department, SSCET Badhani, Pathankot

DOI: 10.5281/zenodo.49716

ABSTRACT

Spell-checking is the process of detecting and correcting incorrectly spelled words in a text paragraph. Spell checking system first detects the incorrect words and then provide the best possible solution of corrected words. Spell checking system is a combination of handcrafted rules of the language for which spell checking system is to be created and a dictionary which contain the accurate spellings of various words. Better rules and large dictionary of words is help to improve the rate of error detection otherwise all the errors cannot be detected. After detecting the wrong or misspelled words ,the various spell correcting techniques are used to provide the best accurate correct words or alternate words which higher the rate of correction of the wrong words. There are many systems available for detecting and correcting text. The system is made to check the spellings and to correct them using various techniques for Punjabi-Hindi text. We used hybrid approach to implement the Spelling checking and Correcting System. This hybrid approach is a combination of “Dictionary look up approach”, “Rule based approach”, “N-Gram Approach”, “Edit Distance approach” and use linguistic features of the Punjabi-Hindi language. This System gives the result accuracy as 91% according to the research work for Punjabi-Hindi words. It gives nearby result up to 91% of words tested in the input data.It gives results for rest of 9% but not the best possible correct word was displayed on the top of the correct word list from the database.

KEYWORDS: Spellchecker, Punjabi-Hindi , Error, dictionary.

INTRODUCTION

Spell-checking is the process of detecting and sometimes providing suggestions for incorrectly spelled words in a text. Spell checking system can be created with the combination of handcrafted rules by considering grammatical features of the language for which spell checking system is to be created and a dictionary which contain the accurate spellings of various words in the target language. Basically, the better the handcrafted rule and larger the dictionary of a spell-checker is, the higher is the error detection rate; otherwise, misspellings would pass undetected. Alas, traditional dictionaries experience from out-of-vocabulary and data sparseness problems as they do not cover large vocabulary of words necessary to cover proper names, domain-specific terms, technical lexica, special acronyms, and terminologies. As a result, spell-checkers will experience less error correction and detection rate and will fail to encounter all errors in the text. All modern mercantile spelling error detection and correction tools work on word level and use a dictionary. Each word from the data or text is searched in the speller lexicon. When a word is not in the dictionary, it is detected as an error. In order to precise the error, a spell checker locates the dictionary for words that look like the wrong word most. These words are then advised to the user who selects the word that was intended. Spelling checking is used in various applications like machine translation, searches, information retrieval etc. There are two main issue related to spell checker. These are error detection and error correction. In developing upon the type of error non word error and real word error. There are many systems available for detecting and correcting text. Spell checker can also be defined as it is a supercomputer application that analysis possible misspelling in a text by referring to the accepted spellings in a database. In the database various accurate words of the target language for which the spell – checker is to be made are stored which consists of proper nouns for males, females, countries, states, rivers, mountains etc. The system is made to check the spellings and to correct them using

various techniques for Punjabi-Hindi text. In this proposed system input in form of a paragraph is given that can include incorrect words and the system will generate the result which contain the accurate text after eliminating the errors. We will use hybrid approach to implement the Spelling checking and Correcting System. This hybrid approach is a combination of “Dictionary look up approach”, “Rule based approach”, “N-Gram Approach”, “Edit Distance approach” and use linguistic features of the Punjabi-Hindi language.

These approaches can be explained in brief as follows:

1.1 Dictionary lookup approach

In this approach each word in the paragraph which will be given as an input is checked for the database entries. If the scanned word is found in the database then is considered to be correct word i.e. spellings of the word are correct but in case the word is not present in the database table then it is considered as an incorrect word. After finding the word incorrect various handcrafted rules are applied to generate the correct spellings of the word by considering the linguistic features of the Punjabi-Hindi language, if approach generate the multiple entries for the single entry then by using statistical analysis a more appropriate word id chosen by the system and is replaced with the incorrect word to generate the result.

1.2 Edit Distance

Edit distance is an easiest technique in spell correction. This easiest method is based on the statement that the person usually makes some mistakes if one, so therefore for each dictionary word the minimum number of the fundamental editing operations (insertion, deletions, substitutions) required to convert a dictionary word in to the non-word .the lower, the number ,the higher the probability that the user has made such errors. Through the operation of adding, deleting and modifying, Edit-Distance changes a word into the minimum operating frequency of another word.

1.3 Rule based Approach

In this approach handcrafted rules are made by considering the features of the Punjabi-Hindi language. These rules are applied on the words in the paragraph which are not found in the database. By the help of these rules the system attempts to generate the exact spellings of the word which is under observation.

N-Gram Analysis

This works when rule based approach fails to generate the appropriate word for the incorrect words. In this approach system try to find the accurate word by considering its neighbor words by comparing with the existing paragraph stored in the system. This method also helps to identify the correct word when more than two words are generated by the rule based approach.

LITERATURE REVIEW

Ritika Mishra, Navjot Kaur, Design and Implementation of Online Punjabi Spell Checker Based on Dynamic Programming, Volume 3, Issue 8, August 2013 ISSN: 2277 128X ,International Journal of Advanced Research in Computer Science and Software Engineering

This paper describes the development and working of online Raftaar Punjabi-Hindi 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-Hindi 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.

Neha Gupta, Pratistha Mathur, Spell Checking Techniques in NLP: A Survey, Volume 2, Issue 12, December 2012 , ISSN: 2277 128X ,International Journal of Advanced Research in Computer Science and Software Engineering

In this paper author describes the various techniques for spell checking and error correction. This paper also provides information about various available spell checking systems developed for various Indian language. In this paper two techniques for spell checking are described which are (1) N Gram Analysis based on statistical technique and (2) is Dictionary lookups. This paper describes the properties of various spell checker and spell Corrector, these systems includes Bangla spell Checker, Oriya Spell Checker , Tamil spell Checker, Marathi spell checker, Punjabi

spell checker etc. Techniques described in this paper for spelling error correction includes "Edit distance", "similarity keys", "Rule Based Techniques", "N-Gram based techniques", "Neural Network based techniques etc.

Baljeet Kaur, Review On Error Detection and Error Correction Techniques in NLP: Volume 4, Issue 6, June 2014 ISSN: 2277 128X , International Journal of Advanced Research in Computer Science and Software Engineering

In this paper we have surveyed the area of spell correction and error detection techniques. Existing work related with spell checkers in Punjabi-Hindi and Punjabi-Hindi language is also discussed. In this paper the author will implement a Punjabi-Hindi spell-checker by using dictionary lookup and edit-distance based technique with more accuracy. In this paper techniques for Error Correction are used (1) N Gram Analysis (2) Rule Based Approach and (3) Edit Distance.

RESEARCH METHODOLOGY

We will use hybrid approach to implement the Spelling checking and Correcting System. This hybrid approach is a combination of "Dictionary look up approach", "Rule based approach", "Statistical Approach, "Edit Distance approach" and use linguistic features of the Punjabi-Hindi language. Dictionary Look up Approach and Edit Distance Approach is used in the research which is already implemented. The system which is to developed will use a hybrid approach to check and to correct the wrong spelled words. Now in this project research I will use the Rule Based Approach and Statistical Approach with more accuracy.

5.1 Following are the steps of proposed algorithm:

Following are the steps of proposed algorithm :

Step I: Input the source string.

Step II: Tokenize the input of first step into words.

Step III For each Token compare it with the Dictionary.

Step IV Check whether it is correct or not. If it is correct, then go to Step III, otherwise apply Rule Bases Approach.

Step V Again find the word from dictionary. If word is found go to Step III, otherwise apply Edit Distance Approach.

Step VI Find the minimum distance from this Token to the word in the Dictionary.

Step VII Sort these words in ascending order of their distance.

Step VIII Check the words obtained with same distance by comparing previous and next word of the target word to obtain best possible suggestion.

Step IX If the combination available in the database then replace the top most word obtained in step VII with token otherwise go to step VII.

Step IX End.



Figure 1.1

5.2 Rule Based Approach

In this approach a set of rules are developed by which the input token is compared. Every rule created in the rule based is applied on the token and corresponding results is produced after applying the rule based approach.

The following are the steps of rule based approach:

Step I: input the Text to find the errors

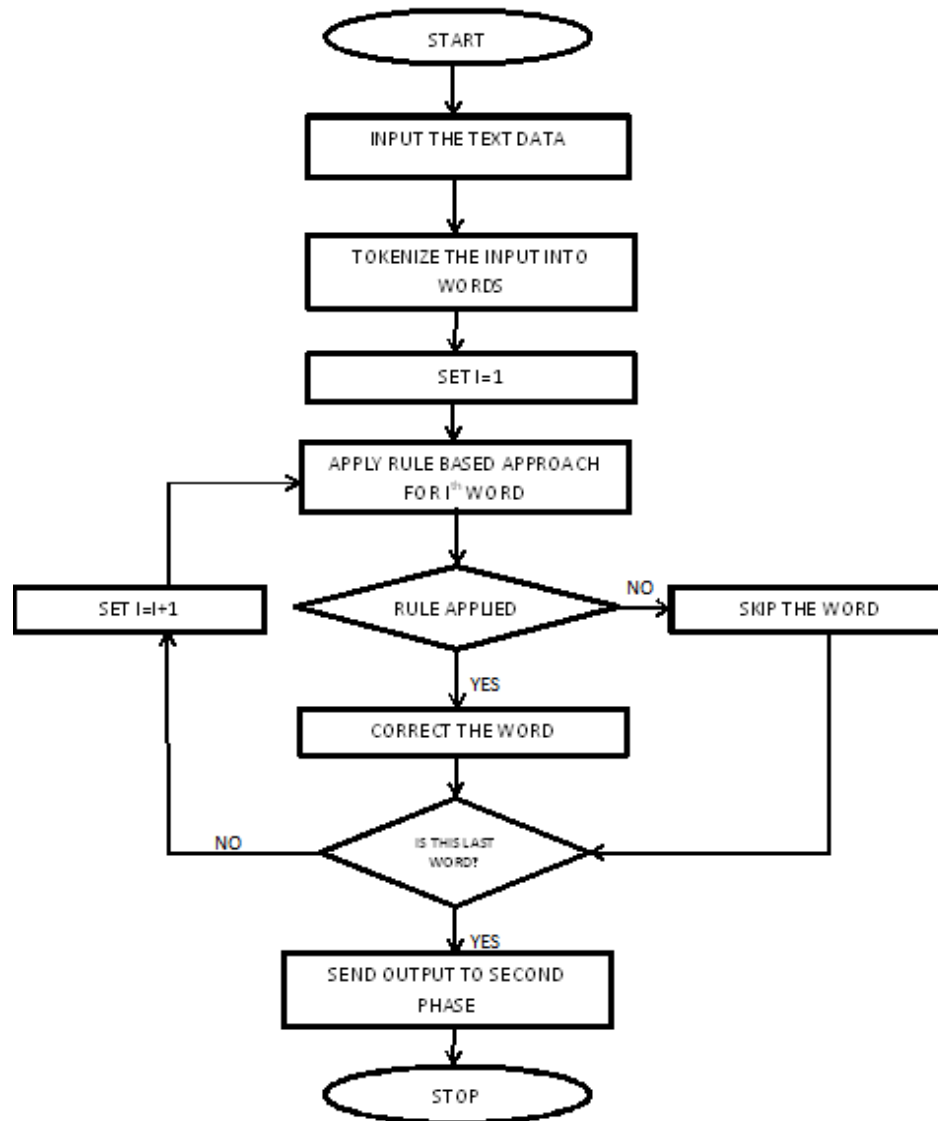
Step II: Tokenize the input of first step into words.

Step III: For each Token compare it with the Dictionary.

Step IV: Check whether it is correct or not. If it is correct, then go to Step III, otherwise apply Rule Based Approach.

Step v: Use output of rule based system into next phase

Step VI: end



FLOW CHART FOR RULE BASED APPROACH

Figure 1.2

5.3 Dictionary lookup technique

This approach is mainly used to check whether the particular token is correct or not by comparing the token with the dictionary values. It is assumed that the word which is being checked is correct if it is available in the dictionary. To create dictionary for various Punjabi-Hindi words, various resources like Punjabi-Hindi text books, online Punjabi-Hindi websites are being used. The accuracy of the system is highly depends upon this phase. If the required word is correct but not in the dictionary then it will give wrong output.

Steps for dictionary look up technique are as follows:

Step I : input the String data to be checked

Step II : Tokenize this input into words

Step III Compare it with the dictionary to check whether it is correct or not.

Step IV End

5.4 Edit Distance Technique

This technique will work if rule based approach becomes unable to generate the accurate word. This technique is used to find the nearest possible word from the dictionary to obtain the result. With the help of this technique various suggestions are generated with respect to the token which is being checked in the ascending order of their distances. In this approach, the word distance means the minimum number of operations required to equate the wrong word with the word in dictionary.

The steps to implement this technique are as follows:

Step I: Input the text string

Step[II: Tokenize the input string

Step III : for each word in the dictionary perform following steps IV and V

Step IV : calculate the distance of word from step III with input token

Step V Store the word and token in the temp location and ignore if distance is more than 3.

Step VI : Sort the words obtained in step V in ascending order and display it to the user.

Step VII End

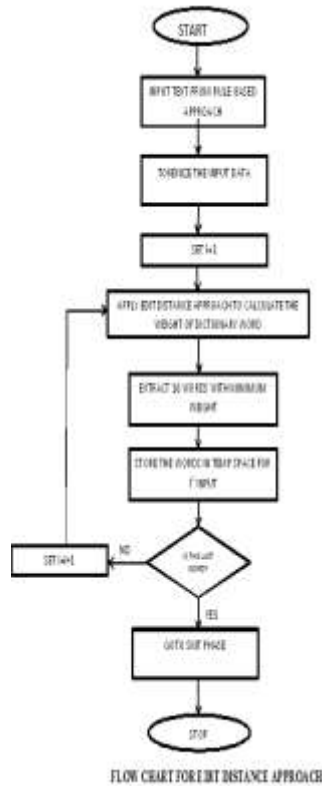


Figure 1.3

RESULT & DISCUSSIONS

To evaluate the performance of the system various inputs from various text books and online websites with some errors in their words are input to the system and results are analyzed. It is calculated that the overall accuracy of the system is 91%.

The result is shown as below:

Input 1:

ਗੁਰੂ ਨਾਨਕ ਦੇਵ ਜੀ ਦਾ ਜਨਮ 15 ਅਪ੍ਰੈਲ 1469[੬] ਨੂੰ ਲਾਹੌਰ, ਪਾਕਿਸਤਾਨ ਨੇੜੇ ਰਾਇ ਭੋਇ ਦੀ ਤਲਵੰਡੀ, ਜਿਸਨੂੰ ਹੁਣ ਨਨਕਾਣਾ ਸਾਹਿਬ ਆਖਿਆ ਜਾਂਦਾ ਹੈ, ਵਿਖੇ ਹੋਇਆ ਜਿਸਨੂੰ ਹੁਣ ਗੁਰੂ ਨਾਨਕ ਦੇਵ ਜੀ ਦੇ ਪ੍ਰਕਾਸ਼ ਦਿਹਾੜੇ (ਗੁਰਪੁਰਬ) ਵਜੋਂ ਮਨਾਇਆ ਜਾਂਦਾ ਹੈ। ਇਸ ਸਥਾਨ 'ਤੇ ਹੁਣ ਗੁਰਦੁਆਰਾ ਜਨਮ-ਅਥਾਨ ਸੁਸ਼ੋਭਤ ਹੈ। ਇਹਨਾਂ ਦੇ ਪਿਤਾ, ਕਲਿਆਣ ਚੰਦ ਦਾਸ ਬੇਦੀ ਜਾਂ ਮਹਿਤਾ ਕਾਲੂ[੭] ਪਿੰਡ ਤਲਵੰਡੀ ਦੇ ਫ਼ਸਲ-ਮਾਮਲੇ ਦੇ ਪਟਵਾਰੀ ਸਨ ਅਤੇ ਉਸ ਇਲਾਕੇ ਦੇ ਇੱਕ ਮੁਸਲਮਾਨ ਜਿਮੀਂਦਾਰ ਰਾਇ ਬੁਲਾਰ ਹੇਠ ਨੌਕਰੀ ਕਰਦੇ ਸਨ।[੮] ਇਹਨਾਂ ਦੀ ਮਾਤਾ ਦਾ ਨਾਂ ਤ੍ਰਿਪਤਾ ਸੀ। ਇਹਨਾਂ ਦੀ ਵੱਡੀ ਭੈਣ ਦਾ ਨਾਂ ਬੇਬੇ ਨਾਨਕੀ ਸੀ ਜੋ ਇਹਨਾਂ ਦੀ ਪਹਿਲੀ ਸਿੱਖ ਬਣੀ।

output

ਗੁਰੂ ਨਾਨਕ ਦੇਵ ਜੀ ਦਾ ਜਨਮ 15 ਅਪ੍ਰੈਲ 1469[੬] ਨੂੰ ਲਾਹੌਰ, ਪਾਕਿਸਤਾਨ ਨੇੜੇ ਰਾਇ ਭੋਇ ਦੀ ਤਲਵੰਡੀ, ਜਿਸਨੂੰ ਹੁਣ ਨਨਕਾਣਾ ਸਾਹਿਬ ਆਖਿਆ ਜਾਂਦਾ ਹੈ, ਵਿਖੇ ਹੋਇਆ ਜਿਸਨੂੰ ਹੁਣ ਗੁਰੂ ਨਾਨਕ ਦੇਵ ਜੀ ਦੇ ਪ੍ਰਕਾਸ਼ ਦਿਹਾੜੇ (ਗੁਰਪੁਰਬ) ਵਜੋਂ ਮਨਾਇਆ ਜਾਂਦਾ ਹੈ। ਇਸ ਸਥਾਨ 'ਤੇ ਹੁਣ ਗੁਰਦੁਆਰਾ ਜਨਮ-ਅਸਥਾਨ ਸੁਸ਼ੋਭਤ ਹੈ। ਇਹਨਾਂ ਦੇ ਪਿਤਾ, ਕਲਿਆਣ ਚੰਦ ਦਾਸ ਬੇਦੀ ਜਾਂ ਮਹਿਤਾ ਕਾਲੂ[੭] ਪਿੰਡ ਤਲਵੰਡੀ ਦੇ ਫ਼ਸਲ-ਮਾਮਲੇ ਦੇ ਪਟਵਾਰੀ ਸਨ ਅਤੇ ਉਸ ਇਲਾਕੇ ਦੇ ਇੱਕ ਮੁਸਲਮਾਨ ਜਿਮੀਂਦਾਰ ਰਾਇ ਬੁਲਾਰ ਹੇਠ ਨੌਕਰੀ ਕਰਦੇ ਸਨ।[੮] ਇਹਨਾਂ ਦੀ ਮਾਤਾ ਦਾ ਨਾਂ ਤ੍ਰਿਪਤਾ ਸੀ। ਇਹਨਾਂ ਦੀ ਵੱਡੀ ਭੈਣ ਦਾ ਨਾਂ ਬੇਬੇ ਨਾਨਕੀ ਸੀ ਜੋ ਇਹਨਾਂ ਦੀ ਪਹਿਲੀ ਸਿੱਖ ਬਣੀ।

Input 2:

ਗੁਰੂ ਜੀ ਵਿਸ਼ਾਲ ਭਾਰਤ ਦੇ ਹਰ ਕੋਨੇ 'ਤੇ ਪਹੁੰਚੇ। ਉਨ੍ਹਾਂ ਮੱਧ ਪੂਰਬ ਵਿਚ ਸਥਿਤ ਇਸਲਾਮੀ ਦੇਸ਼ਾਂ ਦੇ ਲੱਗਭਗ ਸਾਰੇ ਧਰਮ ਕੇਂਦਰਾਂ ਉੱਤੇ ਜਾ ਕੇ ਵਿਵਿਧ ਪ੍ਰਕਾਰ ਦੀਆਂ ਭਾਰਤੀ ਤੇ ਸ਼ਾਮੀ ਧਰਮੀ ਪਰੰਪਰਾਵਾਂ ਨੂੰ ਨੇੜੇ ਹੋ ਕੇ ਵੇਖਿਆ। ਉਨ੍ਹਾਂ ਨੇ ਦਾਰਸ਼ਨਿਕ ਆਧਾਰਾਂ ਤੇ ਅਭਿਆਸ ਪ੍ਰਣਾਲੀਆਂ ਦਾ ਅਧਿਐਨ ਕੀਤਾ ਅਤੇ ਅਗਿਆਨ-ਗ੍ਰਸਤ ਲੋਕਾਂ ਨੂੰ ਕਿਰਿਆਚਾਰੀ ਖੰਡ ਅਤੇ ਰਿੱਧੀਆਂ-ਸਿੱਧੀਆਂ ਅਥਵਾ ਕਰਮਕਾਂਡਾਂ, ਕਰਾਮਾਤਾਂ ਆਦਿ ਦੀ ਨਿਰਾਰਥਕਤਾ ਦੀ ਪ੍ਰੇਰਨਾ ਦਿੱਤੀ। ਆਪ ਜੀ ਦੀ ਲੋਕ ਉਧਾਰਣ ਦੀ ਜੁਗਤੀ ਇਹ ਸੀ ਕਿ ਪ ਕਿਸੇ ਧਰਮ ਉਤਸਵ 'ਤੇ ਇਕੱਤਰ ਹੋਏ ਲੋਕਾਂ ਵਿੱਚ ਜਾ ਕੇ ਅਨੇਖੇ ਨਾਟਕੀ ਅੰਦਾਜ਼ ਨਾਲ ਲੋਕਾਂ ਦਾ ਧਿਆਨ ਆਪਣੇ ਵੱਲ ਖਿੱਚਦੇ ਅਤੇ ਉਨ੍ਹਾਂ ਦੇ ਫੇਕਟ ਕਰਮਾਂ ਦਾ ਖੰਡਨ ਕਰ ਕੇ ਉਨ੍ਹਾਂ ਨੂੰ ਸਤਿ ਵਿਹਾਰ ਕਰਨ ਦੀ ਸਿੱਖਿਆ-ਦੀਖਿਆ ਦੇਂਦੇ। ਅਨੇਕਾਂ ਹੀ ਉਦਾਹਰਣਾਂ ਆਪ ਜੀ ਦੇ ਜੀਵਨ ਵਿੱਚੋਂ ਮਿਲਦੀਆਂ ਹਨ। ਹਰਿਦੁਆਰ ਵਿਖੇ ਸੂਰਜ ਨੂੰ ਪਾਣੀ ਦੇਣ ਦੀ ਬਜਾਇ ਆਪ ਜੀ ਨੇ ਇਸ ਕਰਮ ਦੀ ਨਿਰਾਰਥਕਤਾ ਦਰਸਾਉਣ ਲਈ ਪੱਛਮ ਵੱਲ ਆਪਣੇ ਖੇਤਾਂ ਨੂੰ ਪਾਣੀ ਦੇਣਾ ਸ਼ੁਰੂ ਕੀਤਾ।

Output

ਗੁਰੂ ਜੀ ਵਿਸ਼ਾਲ ਭਾਰਤ ਦੇ ਹਰ ਕੋਨੇ 'ਤੇ ਪਹੁੰਚੇ। ਉਨ੍ਹਾਂ ਮੱਧ ਪੂਰਬ ਵਿਚ ਸਥਿਤ ਇਸਲਾਮੀ ਦੇਸ਼ਾਂ ਦੇ ਲੱਗਭਗ ਸਾਰੇ ਧਰਮ ਕੇਂਦਰਾਂ ਉੱਤੇ ਜਾ ਕੇ ਵਿਵਿਧ ਪ੍ਰਕਾਰ ਦੀਆਂ ਭਾਰਤੀ ਤੇ ਸ਼ਾਮੀ ਧਰਮੀ ਪਰੰਪਰਾਵਾਂ ਨੂੰ ਨੇੜੇ ਹੋ ਕੇ ਵੇਖਿਆ। ਉਨ੍ਹਾਂ ਨੇ ਦਾਰਸ਼ਨਿਕ ਆਧਾਰਾਂ ਤੇ ਅਭਿਆਸ ਪ੍ਰਣਾਲੀਆਂ ਦਾ ਅਧਿਐਨ ਕੀਤਾ ਅਤੇ ਅਗਿਆਨ-ਗ੍ਰਸਤ ਲੋਕਾਂ ਨੂੰ ਕਿਰਿਆਚਾਰੀ ਖੰਡ ਅਤੇ ਰਿੱਧੀਆਂ-ਸਿੱਧੀਆਂ ਅਥਵਾ ਕਰਮਕਾਂਡਾਂ, ਕਰਾਮਾਤਾਂ ਆਦਿ ਦੀ ਨਿਰਾਰਥਕਤਾ ਦੀ ਪ੍ਰੇਰਨਾ ਦਿੱਤੀ। ਆਪ ਜੀ ਦੀ ਲੋਕ ਉਧਾਰਣ ਦੀ ਜੁਗਤੀ ਇਹ ਸੀ ਕਿ ਆਪ ਕਿਸੇ ਧਰਮ ਉਤਸਵ 'ਤੇ ਇਕੱਤਰ ਹੋਏ ਲੋਕਾਂ ਵਿੱਚ ਜਾ ਕੇ ਅਨੇਖੇ ਨਾਟਕੀ ਅੰਦਾਜ਼ ਨਾਲ ਲੋਕਾਂ ਦਾ ਧਿਆਨ ਆਪਣੇ ਵੱਲ ਖਿੱਚਦੇ ਅਤੇ ਉਨ੍ਹਾਂ ਦੇ ਫੇਕਟ ਕਰਮਾਂ ਦਾ ਖੰਡਨ ਕਰ ਕੇ ਉਨ੍ਹਾਂ ਨੂੰ ਸਤਿ ਵਿਹਾਰ ਕਰਨ ਦੀ ਸਿੱਖਿਆ-ਦੀਖਿਆ ਦੇਂਦੇ। ਅਨੇਕਾਂ ਹੀ ਉਦਾਹਰਣਾਂ ਆਪ ਜੀ ਦੇ ਜੀਵਨ ਵਿੱਚੋਂ ਮਿਲਦੀਆਂ ਹਨ। ਹਰਿਦੁਆਰ ਵਿਖੇ ਸੂਰਜ ਨੂੰ ਪਾਣੀ ਦੇਣ ਦੀ ਬਜਾਇ ਆਪ ਜੀ ਨੇ ਇਸ ਕਰਮ ਦੀ ਨਿਰਾਰਥਕਤਾ ਦਰਸਾਉਣ ਲਈ ਪੱਛਮ ਵੱਲ ਆਪਣੇ ਖੇਤਾਂ ਨੂੰ ਪਾਣੀ ਦੇਣਾ ਸ਼ੁਰੂ ਕੀਤਾ।

To evaluate the performance of the system various inputs from various text books and online websites with some errors in their words are input to the system and results are analyzed. It is calculated that the overall accuracy of the system is 91%.

The result is shown as below:

S.No.	Total No. of words in paragraph	Errors in Paragraphs	Correction by Edit Distance Technique	Accuracy of Edit Distance Technique	Errors Corrected by Proposed System	Accuracy of Proposed System
1	75	10	9	90%	10	100%
2	107	15	13	86%	14	93%
3	132	19	16	84%	18	94%
4	150	25	22	88%	24	96%

Table 1.1

Following graph is showing the comparison of existing and proposed system

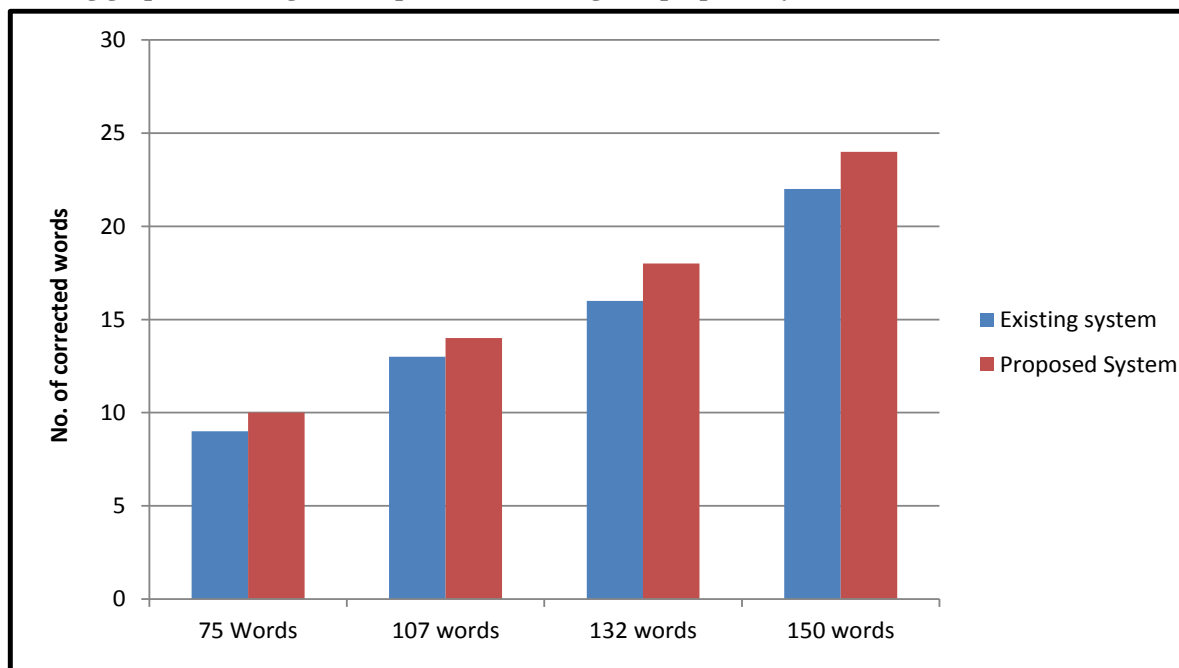


Figure 1.9

Here this graph represents the corrected error by existing system and proposed system. It shows from 75 words system detects 10 error words and existing system corrects 9 and proposed system corrects 10 words. In next input from 107 words there are 15 error words, existing system corrects the 13 words whereas proposed system corrects 14 words and so on.

Accuracy Comparison of existing and proposed system

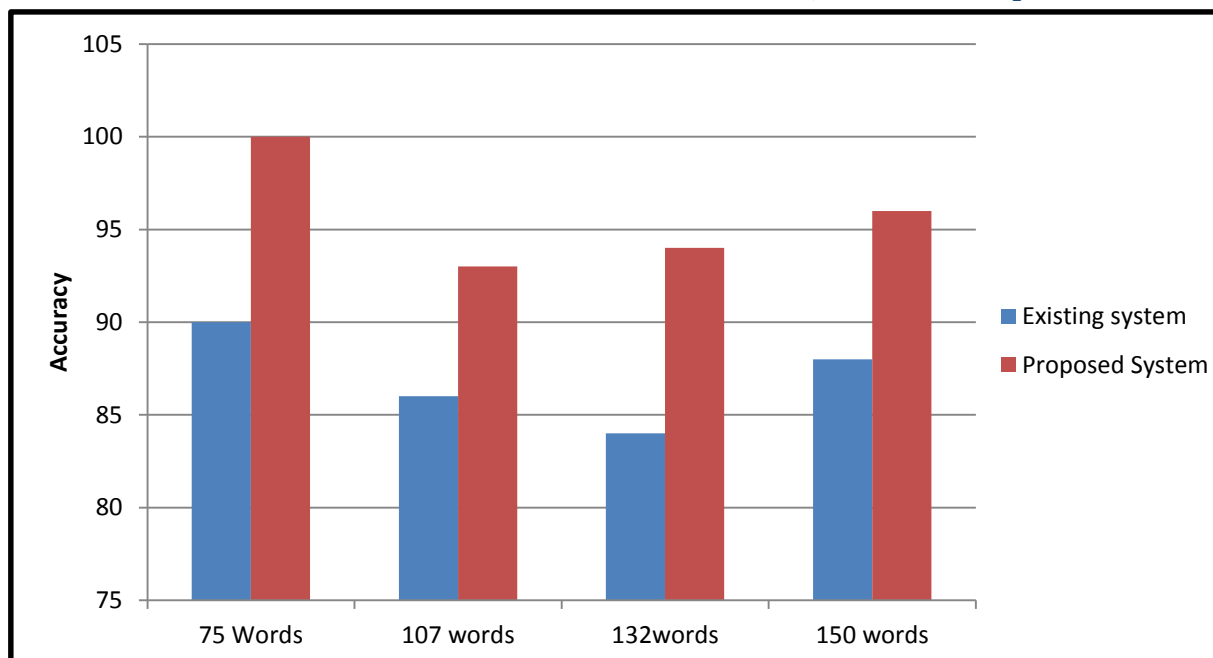


Figure 1.10

This graph shows the accuracy of existing system and proposed system. From the input of 75 words existing system gives 90 % accuracy whereas proposed system gives 100% accuracy and for 107 words existing system gives 86% and proposed system gives 93% accuracy and so on.

CONCLUSION AND FUTURE SCOPE

In our Research work, we have developed an online Punjabi-Hindi spell checker and also developed a new proposed algorithm for the correction of wrong words according to the dictionary. Proposed system is based on hybrid approach in which three approaches which are rule based approach, dictionary look up approach and edit distance approaches are used into one. The main features of Punjabi-Hindi spell checker are large database, online application, easy to operate, email and printing options. This System gives the result accuracy as 91% according to the research work for Punjabi-Hindi words. It gives nearby result up to 91% of words tested in this minor project. It gives results for rest of 9% 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. In future more databases can be added to the system to improve overall accuracy.

REFERENCES

- [1] Ritika Mishra, Navjot Kaur, *Design and Implementation of Online Punjabi Spell Checker Based on Dynamic Programming*, Volume 3, Issue 8, August 2013, ISSN: 2277 128X, International Journal of Advanced Research in Computer Science and Software Engineering
- [2] Neha Gupta, Pratistha Mathur, *Spell Checking Techniques in NLP: A Survey*, Volume 2, Issue 12, December 2012 , ISSN: 2277 128X, International Journal of Advanced Research in Computer Science and Software Engineering
- [3] Baljeet Kaur, *Review On Error Detection and Error Correction Techniques in NLP*: Volume 4, Issue 6, June 2014 ISSN: 2277 128X, International Journal of Advanced Research in Computer Science and Software Engineering.

- [4] 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.
- [5] S. Dasgupta, C.H. Papadimitriou, and U.V. Vazirani, 'Algorithms', p173, available at <http://www.cs.berkeley.edu/~vazirani/algorithms.html>.
- [6] 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.
- [7] Gurpreet Singh Lehal, "Design and Implementation of Punjabi Spell Checker", International Journal of Systemics, Cybematics and Infomatics, 2007.
- [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.
- [16]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.
- [17]Jesus Vilares& Manuel Vilares, "Managing Misspelled Queries in IR Application," Issue 8, October 2010.
- [18]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.