Volume 4, Issue 11, November 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Proposed Technique to Create Online Punjabi Spell Checker Using Unification Method

[1]Er. Sumreet Kaur Randhawa, [2]Er.Charanjiv Singh Saroa

[1]Student, [2]Assistant Professor

Department of Computer Engineering, Punjabi University Patiala, Punjab, India

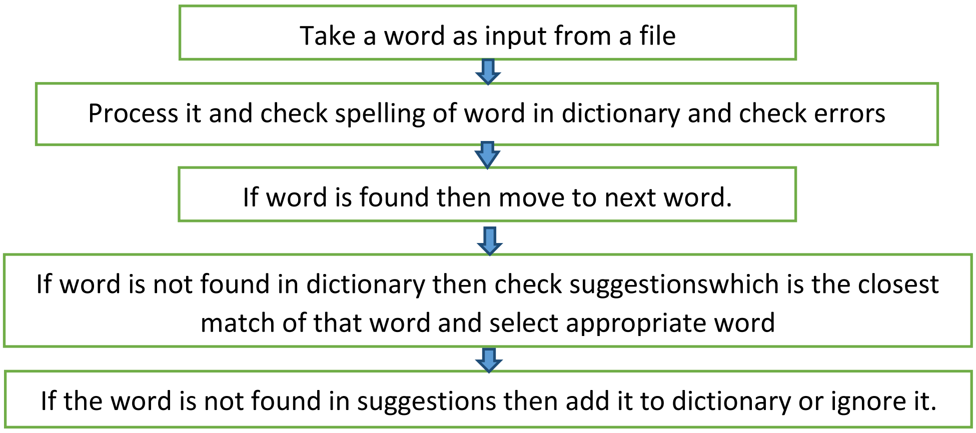
Abstract: Spellcheckers are the basic tools for preparing documentation and for word processing that analyzes possible misspellings in the text by checking it with most accepted spelling in database. Though considerable work has been done in English language but not much work has been done in regional language of India including Punjabi it is a challenge for making spell checker in Punjabi language. It is the world's 12th most widely spoken language.The only available spell checker for Punjabi is AKHAR and SUDHAR. As they are not available online.In this paper we have discussed different techniques and about different spell checker available online in different Indian regional languages. After analyzing all previous techniques we proposed a new technique that will work better than other techniques. In future we will make a Punjabi spell checker using this new technique.

Keywords: Punjabi spell checker, Gurmukhi spell checker,Error detection, Error correction, N-gram,NLP,misspelled words, Types of errors.

I. INTRODUCTION

Spell Checker is an electronic lexicon in a word processor that can be used to hold misspelled words. As we know data are mostly provided in the form of texts, ranging from the reports provided by news portals, using a formal language, to comments in blog and micro-blogging applications that abuse the use of an informal language. Address this heterogeneity is an essential preprocessing so that these data can be used by tools that aim to infer accurate information based on such data. Unfortunately, traditional dictionaries suffer from out-of-vocabulary and data sparseness problems as they do not have large vocabulary of words indispensable to cover proper names, domain-specific terms, technical jargons, etc. As a result, spell-checkers will incur low error detection and correction rate and will fail to flag all errors in the text with the advent of the personal computer, it can be assumed that mistyping of words has increased. Thus, for many of us who do our own typing, the spell checkers in our word processors or other software have become indispensable. Spell Checker has following steps:-

STEPS IN SPELL CHECKER


Take a word as input from a file
Process it and check spelling of word in dictionary and check errors
If word is found then move to next word.
If word is not found in dictionary then check suggestionswhich is the closest match of that word and select appropriate word
If the word is not found in suggestions then add it to dictionary or ignore it.

Error means a measure of the estimated difference between the observed and calculated value Spelling and typing errors are common errors made by humans. Errors may be of missing letters, extra letters, misspelled letters, or disordered letters.In our technique we firstly create dictionary by combining words of dictionary,newspapers, websites, blogs etc. As 40000 words are already available in dictionary and we create corpus of around 7, 00,000 words from which we get around 55,000 unique words. By combining these to make a rich dictionary. We are also surveyed about different online available spell checkers and techniques of spell checkers.

II. LITERATURE SURVEY

As we discussthere are two main issuesrelated to spell checker i.e. error detection and error correction .Further there are two types of errors these are non-word errors and real –word errors or errors may be classified as Typographic errors and Cognitive errors. Many techniques are available for non-word errors. B surveying different spell checkers we came to know about thatIn Malayalam spell checker rule cum dictionary based approach is used, so far the error detection of words is almost complete for any form of agglutination that can come in Malayalam vocabulary. Only standard words are checked in this version and we have further plans to cater the non-standard words too. As far as suggestion generation is considered, only single character errors are taken into account and advanced versions will be taking care of multiple character errors also. In paper"design and implementation of online Punjabi spell checker" raftaar" a proposed algorithm is used for correcting of wrong words. In paper automatic keywords extraction for Punjabi language they include various phrase like removing stop words ,identification of Punjabi nouns etc. which is used for information retrieval, classification, clustering. In paper named "Spell checking techniques in NLP-survey"- describes all techniques of spell checker already available and also spell checkers of different languages. With the help of these papers we came to know about different techniques of spell checker and also about spell checker of different language.

We had made a survey to check that what kind of errors a human being make while writing a word in Punjabi. The error detection process usually consists of checking to see if an input string is a valid index or dictionary word. Efficient techniques have been devised for detecting such types of errors. In our work have selected around200 words of English having different meanings and sounds. These words were dictated to 300 persons to write separately. Then we checked those dictated words to find out the ways a person can misspell those words. From those misspelled words we create a list of common mistakes that a person can make while writing these words and these words are used as a correct word or suggestions. The two most known techniques are n-gram analysis and dictionary lookup. Error correction means just to replace the incorrect with most likely corrected word. Techniques available for error correction are Edit distance, Similarity keys, Rule based technique n-gram based technique, neural technique, Probabilistic technique and neural network. Available websites for spell checkers for different languages are Hindikhoj.com, Star21.com, Shabdkosh.com, khandbahale.com, Shuddhoshabdo.com.

WHY PUNJABI SPELL CHECKER?

Punjabi is an indo Aryan language spoken by almost 110 million native speakers worldwide. It is the third most spoken language in Canada. The language has also significant presence inUnited States of America, Saudi Arabia,and Australia etc. Punjabi which has literary history older than English is 12th most largest spoken language of world has now ultimately came to an end. This is due to absence of local languages in educational system because schools and colleges play a necessary role in preserving languages and culture but now a day'sEnglish language is like punishing language in schools for students and we all know one could better understand things in mother tongue rather than in other language. Although Asian countries like china, japan, are teaching their students in their mother tongue. Suppose in a school science subject is there and rather than understanding concept of science or getting practical knowledge we pay attention to English words used in it. Punjabi speakers in Punjabi is 75,671,704 and in India they are 33,038,280.So Punjabi spell checker act as savior of Punjabi language.

III. TYPES OF ERRORS IN PUNABI LANGUAGE

1. SUBSTITUTION ERRORS (SE): -oneat least one character is substituted by another character this can of mistake is commonly done in Punjabi language. For e.g.:-ਸੁਣਦਾ >- ਬੁਣਦਾ, ਭਾਸ਼ਾ>-ਬਾਸ਼ਾ

2. DELETIONERROR (DE)-: when one character is deleted which is although present in desired word. For e.g.:-ਜੰਮੂ>-ਜਮੂ, ਵਿਸ਼ਾਲ>-ਵਿਸ਼ਾ,

3. INSERATIONERRORS (IE):-when one extra character is added in desired word for e.g.:-.ਜਾਣਕਾਰੀ>-ਜਾਣਾਕਾਰੀ, ਸਾਰ>-ਸਾਰਾ

4. RUN-ON-ERROR (ROE):-whentwo words are mistakenly written side by side there is no space between two or more valid words. For e.g.ਜਿਸਦਾ>- ਜਿਸ ਦਾ, ਕਰਦਾ>-ਕਰ ਦਾ

5. TRANSPOSITIONERRORS (TE):-When two adjacent character of words are typed in swapped manner. For e.g.ਰਾਤ>-ਰਤਾ, ਕਰਦਾ>-ਕਦਰਾ

6. SPLIT WORD ERRORS (SWE):-when an extra space is added between a word or two words. For e.g.ਜਿਸਦੇ>-ਜਿਸ ਦੇ, ਉਸਦੇ>-ਉਸ ਦੇ

IV. AVAILBLE TECHNIQUES OF SPELL CHECKER ERROR DETECTION TECHNIQUE

1. N GRAM ANAYLSIS

It is used for finding incorrect word in the text. Instead of comparing each entire word in a text to a dictionary, just n-grams are controlled. A check is done by using an n-dimensional matrix where real n-gram frequencies are stored. If a non-existent or rare n-gram is found the word is flagged as a misspelling, otherwise not. An n-gram is a set of consecutive characters taken from a string with a length of whatever n is set to. If n is set to one then the term used is a unigram, if n is two then the term is a Bigram, if n is three then the term is trigram. N-gram tables can take on a variety of forms but the simplest is bi-gram array which is 2-D array of size 41*41 whose element represents all possible 2-D letter combination of alphabet. The n-grams algorithms have the major advantage that they require no knowledge of the language that it is used with and so it is often called language independent or a neutral string matching algorithm. Using n-grams to calculate for example the similarity between two strings is achieved by discovering the number of unique n-grams that they share and then calculating a similarity coefficient, which is the number of the n-grams in common (intersection), divided by the total number of n-grams in the two words (union)

2. DICTIONARY LOOKUP

This technique simply lookup every word in the dictionary if the word is not there then it is said to be an error. Dictionaries have their own characteristics and storage requirements. It is a straightforward task. Large dictionary might be a dictionary with most common word combined with a set of additional dictionaries for specific topics such as computer science or economy .Big dictionary also uses more space and may take longer time to search. The non-word errors can be detected as mentioned above by checking each word against a dictionary. The drawbacks of this method are difficulties in keeping such a dictionary up to date. At the same time one should keep down system response time. Too small a dictionary can give the user too many false rejections of valid words. The most common technique used for gaining fast access in dictionary is Hash tables. In order to lookup a string, one has to compute its hash address and retrieve the word stored at that address in the pre constructed hash table. If the word stored at the hash address is different from the Input string, a misspelling is flagged. Hash tables main advantage is their random-access nature that eliminated the large number ofcomparisons needed to search the dictionary. The main disadvantage is the need to devise a clever hash function that avoids collisions. To store a word in the dictionary we calculate each hash function for the word and set the vector entries corresponding to the calculated values to true.

ERROR CORRECTION TECHNIQUES

1. EDIT DISTANCE

Edit Distance Edit distance is a simple technique first edit distance spelling error correction algorithm was implemented by Damerau Simplest method is based on the assumption that the person usually makes few errors if ones, i.e. errors from keyboard input therefore for each dictionary word .The minimal number of the basic editing operations (insertion, deletions, substitutions) necessary to covert a dictionary word in to the non-word. It is not quite as good for correcting phonetic spelling errors.

2. N-GRAM TECHNIQUE

N-grams can be used in two ways, either without a dictionary or together with a dictionary. Letter N-gram including tri-gram, bi-gram and uni- gram have been used in variety of ways in text recognition and spelling correction techniques. Used without a dictionary, n-grams are employed to find in which position in the misspelled word the error occurs. If there is a unique way to change the misspelled word so that it contains only valid n-grams, this is taken as the correction. The performance of this method is limited.

3. SIMILARITY KEYS

In this technique we map every string into a key such that the similarly spelled strings will have similar key. It is known as SOUNDEX system. In this it is not necessary to directly compare misspelled string to each word in dictionary. For example suppose a customer comes in a bank and said his name is zayheijendn. So in this case you cannot ask him to speak his name as his English is poor and others customers are waiting. So we want a key which sounds like his name and find a name resembles with it.

4. RULE BASED TECHNIQUE

This methods are interesting approach used in all spell checkers. Edit distance can be viewed as a special case of a rule based technique.

ਗਲਤ ਸ਼ਬਦ	ਸਹੀ ਸ਼ਬਦ	ਨਿਯਮ
AIdq	Awdq	A Sbd nwl “I” nhI l~gdI
aUwT	ਊਠ	a Sbd nwl “w” nhI l~gdw

5. NEURAL NETWORK

This method works on small dictionaries. Back propagation algorithm is used in neural network.it consist of 3 layers input, hidden and output layer. In this input information is represented by on- off pattern. A=1 indicates that node is turned on and A=0 means node is turned off. For e.g. in spell checking applications a misspellings represented as binary n-gram vector may be taken as input and output pattern might be vector of m elements means number of words in lexicon.

6. PROBALISTICS TECHNIQUES

Based on some statistical features of the language-gram technique led to probalistic technique in spell correction and text recognition. Two methods are used in this. Transition probabilities which is similar to n-grams .It is language indepdent. Confusion probabilities estimates of how often a given letter is mistaken. It is source indepdent.

V. AVAILABLE ONLINE SPELL CHECKERS

There are many spell checkers for Indian languages are developed by using above techniques. This section provides brief discussion of some available spell checkers and websites available for those spell checkers and their ranking on 5 November 2014. The number of spell checker available for different regional languages are Hindi, Marathi, Bangla, Tamil, Malayalam, and Punjabi spell checkers.

Table 2

Name of spell checker	Website	Global rank	India rank	Daily page viewer per visitor	Total sites linking in
Hindi spell checker	SHABDKOSH.COM	3,497	296	4.62	462
Marathi spell checker	KHANDBAHALLE.COM	320,723	30,873	2.50	10
Bangla spell checker	SHUDDHOSHABDO.COM	19,779,445	N/A	1	3

VI. PUNJABI SPELL CHECKER

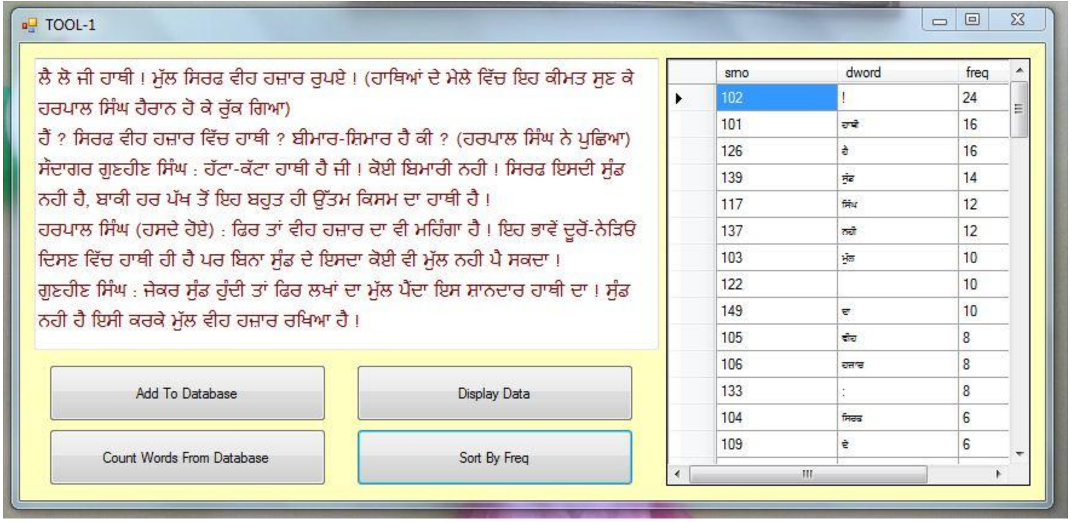
Akhar (Punjabi Spell Checker)

It is an offline spell checker. It is a language sensitive Punjabi / English spell checker has been provided in Akhar. Akhar can automatically detect the language and invokes the respective spell checker. The Unicode complaint Punjabi Spell Checker is font independent and can work on any types of the popular Punjabi fonts such as, Anantpur Sahib, Amritlipi, Jasmine, Punjabi, Satluj etc. This removes the contrast on the user to type the text in pre-defined font only.

VII. PROPROSED TECHNIQUE FOR PUNJABI SPELL CHECKER

1.Firstly we have created a dictionary by combining words of dictionary, newspapers, websites, blogs etc. As 40000 words are already available in dictionary and we create corpus of around 7, 00,000 words from which we get around 55,000 unique words. By combining these we got rich dictionary.

TOOL TO SAVE WORDS ACCORDING TO FREQUENCY IN DATABASE



ਲੈ ਲੋ ਜੀ ਹਾਥੀ ! ਮੁੱਲ ਸਿਰਫ ਵੀਹ ਹਜ਼ਾਰ ਰੁਪਏ ! (ਹਾਥਿਆਂ ਦੇ ਮੇਲੇ ਵਿੱਚ ਇਹ ਕੀਮਤ ਸੁਣ ਕੇ ਹਰਪਾਲ ਸਿੰਘ ਹੈਰਾਨ ਹੋ ਕੇ ਰੁੱਕ ਗਿਆ)
ਹੈਂ ? ਸਿਰਫ ਵੀਹ ਹਜ਼ਾਰ ਵਿੱਚ ਹਾਥੀ ? ਬੀਮਾਰ-ਸ਼ਿਮਾਰ ਹੈ ਕੀ ? (ਹਰਪਾਲ ਸਿੰਘ ਨੇ ਪੁਛਿਆ)
ਸੌਦਾਗਰ ਗੁਣਹੀਣ ਸਿੰਘ : ਹੱਟਾ-ਕੱਟਾ ਹਾਥੀ ਹੈ ਜੀ ! ਕੋਈ ਬਿਮਾਰੀ ਨਹੀ ! ਸਿਰਫ ਇਸਦੀ ਸੁੰਡ ਨਹੀ ਹੈ, ਬਾਕੀ ਹਰ ਪੱਖ ਤੋਂ ਇਹ ਬਹੁਤ ਹੀ ਉੱਤਮ ਕਿਸਮ ਦਾ ਹਾਥੀ ਹੈ !
ਹਰਪਾਲ ਸਿੰਘ (ਹਸਦੇ ਹੋਏ) : ਫਿਰ ਤਾਂ ਵੀਹ ਹਜ਼ਾਰ ਦਾ ਵੀ ਮਹਿੰਗਾ ਹੈ ! ਇਹ ਭਾਵੇਂ ਦੂਰੋਂ-ਨੇੜਿਓ ਦਿਸਣ ਵਿੱਚ ਹਾਥੀ ਹੀ ਹੈ ਪਰ ਬਿਨਾ ਸੁੰਡ ਦੇ ਇਸਦਾ ਕੋਈ ਵੀ ਮੁੱਲ ਨਹੀ ਪੈ ਸਕਦਾ !
ਗੁਣਹੀਣ ਸਿੰਘ : ਜੇਕਰ ਸੁੰਡ ਹੁੰਦੀ ਤਾਂ ਫਿਰ ਲਖਾਂ ਦਾ ਮੁੱਲ ਪੈਂਦਾ ਇਸ ਸ਼ਾਨਦਾਰ ਹਾਥੀ ਦਾ ! ਸੁੰਡ ਨਹੀ ਹੈ ਇਸੀ ਕਰਕੇ ਮੁੱਲ ਵੀਹ ਹਜ਼ਾਰ ਰਖਿਆ ਹੈ !

smo	dword	freq
102	!	24
101	ਹਾਥੀ	16
126	ਹੈ	16
139	ਸੁੰਡ	14
117	ਸਿੰਘ	12
137	ਨਹੀ	12
103	ਮੁੱਲ	10
122		10
149	ਦਾ	10
105	ਵੀਹ	8
106	ਹਜ਼ਾਰ	8
133	:	8
104	ਸਿਰਫ	6
109	ਦੇ	6

2. Secondly we have selected around 200 words of Punjabi having different meanings and sounds. These words were dictated to around 300 persons to write separately. Then we checked those dictated words to find out the ways a person can misspell those words and find errors as shown below:-

ROUND1:-Words having highest frequency error rate.

ROUND2:-Words having low frequency error rate as compared to round 1.

ROUND3:- Words having low frequency error rate means these errors are rarely made by humans.

Table 3

s.no	Right word	Round 1	Round 2	Round 3
1	ਯੂਨੀਵਰਸਿਟੀ	ਯੂਨੀਵਰਸੀਟੀ	ਯੂਨੀਵਰਸਟੀ	ਯੂਨਿਵਰਸਿਟੀ
2	ਪੇਸ਼	ਪੇਸ	ਪੇਛ	ਪੈਸ਼
3	ਭਾਸ਼ਾ	ਭਾਸਾ	ਪਾਸਾ	ਬਾਸਾ
4	ਸਭਿਆਚਾਰਕ	ਸਬਿਆਚਾਰਕ	ਸਭਆਚਾਰਕ	ਸਭੀਆਚਾਰਕ

5	ਵਡਮੁੱਲੇ	ਵਡਮੁਲੇ	ਵਦਮੁੱਲੇ	ਵਦਮੂਲੇ
6	ਪ੍ਰੋਗਰਾਮਾਂ	ਪ੍ਰੋਗਰਾਮਾ	ਪੋਗਰਾਮਾ	ਪ੍ਰੋਗਰਾਮਾ
7	ਲਾਜ਼ਮੀ	ਲਾਜਮੀ	ਲਜਮੀ	ਲਾਜਮਿ
8	ਮੁਥਾਜ	ਮਥਾਜ	ਮੁਹਥਾਜ	ਮੁਹਥਾਜ
9	ਪ੍ਰਤਿਭਾ	ਪ੍ਰਤੀਬਾ	ਪ੍ਰਤਿਭਾ	ਪਤਿੲਾ
10	ਫ਼ਾਰਸੀ	ਫਾਰਸੀ	ਫਾੜਸੀ	ਫਾਰਸੀ
11	ਮੁਸ਼ਕਲ	ਮੁਸਕਲ	ਮੁਸਕਿਲ	ਸੁਸਕੀਲ
12	ਗ੍ਰੰਥ	ਗੰਰਥ	ਗ੍ਰਥ	ਗਰੰਥ
13	ਰੱਖਦਿਆਂ	ਰੱਖਦਿਆ	ਰੱਖਦੀਆ	ਰਖਦੀਆ
14	ਵਿਦਵਾਨ	ਵਿਦਵਾਣ	ਵਿਧਵਾਸ	ਵੀਦਵਾਸ
15	ਸੁਆਗਤ	ਸਆਗਤ	ਸੁੲਾਗਤ	ਸਵਾਗਤ
16	ਸੁਚੱਜੀ	ਸਚੱਜੀ	ਸੁਚਜੀ	ਸਚਜੀ
17	ਰਚਨਾਵਾਂ	ਰਚਨਾਵਾ	ਰਚਿਨਾਵਾ	ਰਚੀਨਾਵਾ
18	ਅਨੁਵਾਦ	ਅਨਵਾਦ	ਅਣਵਾਦ	ਅਨਿਵਾਦ
19	ਸਰੲੋਤਮ	ਸਰਵੋਤਮ	ਸਰੲੈਤਮ	ਸਰਬੌਤਮ
20	ਮਨਮੁਖ	ਮਨਮਖ	ਮਣਮੁਖ	ਮਨਸੂਖ
21	ਦ੍ਰਿਸ਼ਟੀ	ਦ੍ਰਿਸਟੀ	ਦ੍ਰੀਸ਼ਟੀ	ਦਿਰਸਟੀ
22	ਸ੍ਰੋਤ	ਸਰੋਤ	ਛਰੋਤ	ਸਰੌਤ
23	ਖੰਸਿਆ	ਪੰਸੀਆ	ਪੰਸਆ	ਪਸਿਆ
24	ਫ਼ੋਟੋ	ਫੋਟੋ	ਫੁਟੋ	ਫੌਟੋ
25	ਵਿਸ਼ਾਲ	ਵਿਸਾਅ	ਵੀਸਾਲ	ਵੀਛਾਲ
26	ਪਾਠਕਾਂ	ਪਾਠਕਾ	ਪਾਧਕਾ	ਪਾਥਕਾ
27	ਅਣਥੱਕ	ਅਸਥਕ	ਅਣਥਕ	ਅੱਣਥਕ
28	ਪ੍ਰੇਰਨਾ	ਪੇਰਨਾ	ਪੇਰਣਾ	ਪਰੇਣਾ
29	ਸਖ਼ਸੀਅਤ	ਸਖਸੀਅਤ	ਸਕਸੀਅਤ	ਸਖਸੀਅਤ
30	ਕਰਨਾਟਕ	ਕਰਣਾਟਕ	ਕਰਨਟਕ	ਕਰਣਾਟਕ
31	ਪੀਡੀਆ	ਪਿਡੀਆ	ਪੀਡਿਆ	ਡੀਡੀਆ
32	ਯੂਨੀਅਨ	ਜੁਨੀਅਨ	ਜੁਨੀਅਣ	ਜੁਨੀਅਣ

3. We know that error is a mistake or measure of estimated difference between observed and calculated value. Now from Table 3we came to know about what kind of mistakes usually humans make:-

Table 4

s.no	RightChar -> WrongChar	RightChar ->Wrong char
1	ੀ -> ਿ	ਿ -> ੀ
2	ੁ ->ੂ	ੂ ->ੁ
3	ਭ ->ਬ	ਬ ->ਭ
4	ਲ਼ ->ਲ	ਲ ->ਲ਼
5	ਸ਼ ->ਸ	ਸ ->ਸ਼
6	ਣ ->ਨ	ਨ ->ਣ
7	ਊ ->ੳ	ੳ ->ਊ

8	ਜ ->ਝ	ਝ ->ਜ
9	ਦ ->ਧ	ਧ ->ਦ
10	◌ੇ ->◌ੋ	◌ੋ ->◌ੇ
11	ਜ ->ਝ	ਝ ->ਜ
12	◌ੇ ->◌ੈ	◌ੈ ->◌ੇ

4. With the help of this we are able to give appropriate suggestions for misspelled word.

VIII. CONCLUSION AND FUTURE WORK

In this paper wehave analyze type of errors made by humans in Punjabi languageand we combine words of dictionary and newspaper and save them in our database according to frequency to make rich dictionary.We had done survey on the spell checker techniques and websites of available spell checker in regional language. We have also discussed about importance of Punjabi language .In future we will make a Punjabi spell checker with the help of these techniques with efficient database.

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] Sumreet Kaur Randhawa and charanjiv Singh Saroa, “ study of spell checker techniques and available spell checker in regional languages,” International journal for Technical Research in Engineering, volume 2,issue 3 , November-2014.

[3] 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, Cybemetics 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. Vol. 5, No.3, May 2012]

[7] Daniel Jurafsky, James H. Martin, Speech and Language Processing, PEARSON, 2nd ed. [2] Dr.T.V Geeta, Dr.Rajani Parthearath, “Tamil spellchecker”, Resource center for Indian language Technology Solution, TDIL newsletter

[8] Dr. R.K Sharma, “The Bilingual Punjabi English spell checker,” Resource center for Indian language Technology Solution, TDIL newsletter

[9] “F.J Damerau (1964)”A technique for computer detection and correction of spelling error”, communication ACM.

[10] Http:// www.Baraha.com[11] Manisha Das, S.Borgohain, Juli Gogai, S.B Nair (2002),”Design and implementation of a spell checkers for Assamese”, Language Engineering Conference.

[12] Mukand Roy, Gaur Mohan, Karunes K arora,”Compartive study of spell checker algorithm for building a generic spell checkers For Indian language C-DAC NODIA, India.

[13] Malayalam spell checker, Santhosh. T. Varghese, K.G. Sulochana, R. Ravindra Kuma

[14] Alexa.com