Analyzing the Punjabi Language Stemmers: A Critical Approach

Harjit Singh

APS Neighbourhood Campus, Punjabi University, Patiala, India

Abstract

Stemming is a procedure used to reduce the inflected words by removing affixes from them. It is widely used in Natural Language Processing systems developed for various languages like English, Hindi etc. Punjabi is a low resource availability language spoken in northern regions of India and Pakistan. In Pakistan, Shahmukhi script is used to write Punjabi. While in India, Gurmukhi script is used. Various approaches are used for stemming natural language words such as Brute Force approach, Rule based approach, Statistical approach etc. In Gurmukhi scripted Punjabi language stemming, either the Brute Force approach or Rule based approach or a combination of both of these approaches have been used till now. This paper presents a comparative analysis of various stemmers developed so far for stemming Gurmukhi scripted Punjabi language words based on their methodology and accuracy. It will motivate further research in developing more efficient stemmers for Punjabi language. It will benefit the researchers to compare and understand the used approaches and the problems faced in adopting each approach.

Keywords

Stemming, Natural Language Processing, Information Retrieval, Suffix Striping, Punjabi Language Processing

1. Introduction

A procedure adopted to reduce the inflected words by removing any affixes attached to them is called Stemming. Stemming is a very common and crucial task performed in Natural Language Processing (NLP) applications. Many morphologically similar words can be stemmed to the same stem word after removing any suffixes or prefixes from them. It improves the efficiency of further processes adopted for information retrieval. It improves the effectiveness of information retrieval by helping to identify redundant information, to remove irrelevant information from retrieval and to optimize the retrieval by placing high ranked information on the top [1].

Various approaches are used for stemming natural language words such as Brute Force approach, Rule based approach, Statistical approach etc. In brute force approach, a table of inflected words along with their stem words is maintained. To stem a word, that word is searched in the table for a match. If match occurs with some word, the related stem word is picked from the table. The approach is easy to implement and the accuracy is dependent on the table of words available in the database. In rule based approach, the affixes are removed by using some fixed rules. The input words are checked to have a group of characters at the end or beginning of each word. That set of characters (affix) is removed to get the stem word. In some words, an alternative set of

characters is attached to the word to complete it.

In statistical approach, the distance measure is calculated between the two words. If the words are morphologically similar, they will have low distance measure. For morphologically unrelated words, distance measure will be high. Using this measure, the words of a document are grouped to make clusters. In this way, morphologically similar words will be grouped in each cluster. Then the central word of each cluster is taken as a stem word [2].

In Gurmukhi scripted Punjabi language stemming, either the Brute Force approach or Rule based approach or a combination of both of these approaches have been used till now. This paper analyses all those approaches and compares them with each other based on the methodology and accuracy. It will motivate further research in developing more efficient stemmers for Punjabi language. It will benefit the researchers to compare and understand the used approaches and the problems faced in adopting each approach.

1.1. Punjabi Language

Punjabi language belongs to Indo-Aryan group of languages. Its word order in sentence is Subject-Object-Verb and is spoken by almost 100 million people residing in India, Pakistan and many other countries. It may be called Punjabi community. Table 1 shows the number of Punjabi speakers in some countries.

Table 1
Punjabi Speakers in Some Countries

Country	Punjabi Speakers	Source
Pakistan	99,774,008	www.cia.gov
India	33,124,726	censusindia.gov.in
Canada	543,495	www12.statcan.gc.ca
UK	273,000	www.ons.gov.uk
US	253,740	www.census.gov
UAE	191,000	ethnologue.com

Punjabi is basically written in two different scripts; Perso-Arabic and Gurmukhi. Perso-Arabic is related to Persian script which is also called Shahmukhi, and Gurmukhi is derived from Landa, an old script of Indian sub-continent.

2. Literature Review

Jain et al. [3] used NLP technique to find root word in Hindi language. The Hindi inflected word was stemmed to remove affix from the end of the word. The stemmer was tested using 800 inflected words. They used prefix, suffix and root word lists to remove the affixes and validate the root word. The stemmer provided the hit ratio of 0.93. Yadav et al. [4] proposed a method for automatic summarization of text documents. They generated a network of related sentences based on semantic and lexical relation. Two sentences are related to each other using a certain strength that represents the number of relationships between the two. Gros et al. [5] performed an analysis to test the effect on the performance of information retrieval with anaphora resolution. They claimed the positive effect on the results. In this way, the quality of information retrieval results can be significantly improved.

Agrawal et al. [6] proposed a translation tool to translate Sanskrit language text to Hindi. They used stemming in preprocessing the Sanskrit words before converting them to Hindi. It was used as a major phase in translation tool development. Any affixes from input Sanskrit words were removed before using them for generating the equivalent Hindi text. A Sanskrit corpus was used to successfully stem the words by searching them in the corpus and validate.

Bhadwal et al. [7] developed Hindi-Sanskrit bilingual translation system. The system was able to translate Hindi text to Sanskrit and vice-versa using a number of modules. For Sanskrit to Hindi translation flow, they used stemming as a major step before syntax analysis. The system was developed in Java. Bhadwal et al. [8] proposed a Hindi to Sanskrit machine translation system. The system includes a number of NLP modules such as transliteration, POS-tagging, root verb extraction etc. In root verb extraction, the algorithm finds the root verb before mapping to equivalent Sanskrit word.

Although there are positive effects of stemming in various NLP applications, but Wahbeh et al. [9] claimed negative effects on text classification in Arabic. They used stemming as a preprocessing module in Arabic

text classification and found that results were negatively affected when stemming was used.

Review of literature shows that stemming is used for a variety of NLP applications. It motivated the analysis of stemmers developed for Punjabi language. Only a handful of research papers found for this low resource availability Asian language.

3. Stemming Linguistics in Gurmukhi Punjabi

In linguistics, an affix refers to either suffix or prefix. In Punjabi language, both types of affixes are used to make proper words for sentences. But in most cases, prefix changes the meaning of the word. So, in NLP applications, only suffixes are considered for removal. Table 2 shows some inflected words with their stem words. The table shows how the Punjabi words are dealt with linguistically to remove some characters and if required, substitute some alternative set of characters to complete the word.

Table 2
Some Punjabi Inflected and Root words

Inflected Word	Stem word	Suffix Removal and Substitution
ਵਿਦਿਆਰਥੀਆਂ	ਵਿਦਿਆਰਥੀ	ਆਂ
ਬੱਸਾਂ	ਬੱਸ	ਾਂ
ਕੁੱਤੇ	ਕੁੱਤਾ	ੇ and ਾ
ਹਵਾਵਾਂ	ਹਵਾ	ਵਾਂ
ਘਰੋ	ਘਰ	ੋ
ਮੁੜੇ	ਮੁੜ	ੇ
ਕੁੱਤਿਆ	ਕੁੱਤਾ	ਿਆ and ਾ
ਖੇਤੋ	ਖੇਤ	ੋ
ਥੱਲਿਓ	ਥੱਲੇ	ਿਓ and ੇ

4. Analysis of Gurmukhi Punjabi Stemmers

Regardless of the language, stemming is an NLP process to find root words. So, this paper is understandable to the readers even without the knowledge of Punjabi and Gurmukhi script, because function of a stemmer is just to remove any affixes attached to a word. Therefore only brief introduction to Punjabi and Gurmukhi script is enough. This paper discusses the stemmers developed for Punjabi language Gurmukhi script. The paper discusses the methodologies used by researchers to stem the inflected words. In this paper, Gurmukhi scripted Punjabi language henceforth will be called Punjabi.

The significance of this analysis is that it discloses the gap in the NLP research for Punjabi language. It will motivate further research in developing more efficient stemmers for Punjabi using new ideas or by extending existing approaches.

4.1. Earlier Stemmer for Punjabi

Kumar et al. [10] proposed a Punjabi stemmer based on combination of brute force approach and suffix stripping technique. The researchers divided the system in three units and named these units as Input unit, Output unit and Processing unit.

The Input unit takes a Punjabi word and that word is sent to the Processing unit which searches for the word in a database of Punjabi inflected words and their stem words. If a word is found in the database, the corresponding stem word is considered as output stem word. If no match is found in database, then suffix removal technique is used to stem the word. The word ending is searched in a list of possible Punjabi suffixes. If the word ending matches with any one of the suffix present in the suffix list then the ending (suffix) is removed to get the stem word. The steps are shown in Algorithm 1 and graphically represented in Figure 1.

Algorithm 1: Earlier Stemmer for Punjabi
1: Input Punjabi Word for stemming.
2: Search the word in database.
3: If Found; goto (6).
4: Search the suffix of word in suffix list.
5: If Found; remove the suffix.
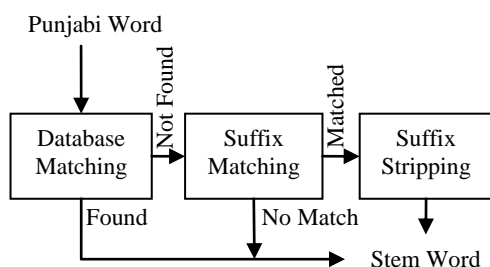*6: Display stem word.*
7: End

Figure 1: Earlier Stemmer for Punjabi

Proposed stemmer was evaluated using three parameters i.e. Correctness, Effectiveness and Performance. Its accuracy is dependent on the number of words stored in the database. Researchers stored 28000 words in the database with 7100 root words for Brute Force search to find matching root word at the first step. So it reduced the chances of using the second step of removing suffix. More the number of words in the database more will be the correctness of the system. For the database they took words from a Punjabi-English dictionary and some online blogging and news websites.

Effectiveness of the system is determined by the behavior of the system in abnormal conditions, such as entering a word that does not exist. This stemmer simply displays the same word as the output. The problems of over-stemming and under-stemming can be further reduced by providing the large database of words. But if the word is not available in the database, these problems may arise in the second step i.e. removing suffix from the inflected word.

Performance of the system is determined by its output accuracy. This system is totally dependent on its database of 28000 words with 7100 root words. The accuracy was calculated by dividing the number of correct outputs by the total number of inputs. The calculated average accuracy of the system was 80.73%.

4.2. Improved Version of Earlier Stemmer

Kumar et al. [11] proposed an improved version of Punjabi stemmer based on Brute force search with suffix stripping technique. In this version of stemmer a database is prepared containing inflected words along with their stem words. The stemmer tries to match the Punjabi word with words stored in the

database. On finding a match, the stem word retrieved from the database is given as output. Algorithm 2 shows the working of the stemmer. Graphically representation is shown in Figure 2.

Algorithm 2: Improved Version of Earlier Stemmer
1: Input Punjabi word for stemming.
2: Search the Word in Exception List.
3: If Found; goto (9).
4: Search the word in database.
5: If Found; goto (8).
6: Search the suffix of word in suffix list.
7: If Found; remove the suffix.
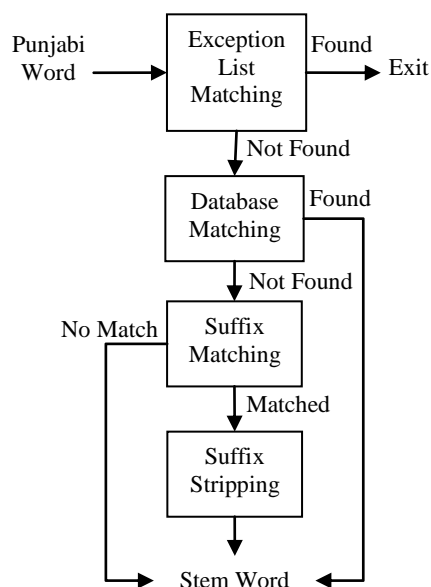*8: Display stem word.*
9: End

Figure 2: Second Version of Earlier Stemmer

The proposed stemmer was evaluated for correctness, effectiveness and its performance. This stemmer used a database of 52000 words from various fields which contained 14400 root words. So it decreased the number of chances to execute the procedure of removing suffix from the inflected word. Increasing the number of words in the database will result in improved correctness of the output. Words were added to the database from various sources such as www.punjabionlinedictionary.com, Jag Bani Newspaper, Pardeep Punjabi to English Dictionary and some online blogging and news websites.

To measure the effectiveness of the stemmer, its behavior was verified in unusual conditions. For invalid input word, this stemmer provides the same word as output. A non-existing word in the database may result in over or under-stemming at suffix removal process. To overcome this problem, an exception list of approximately 200 words was created. These are those words that cause over-stemming and under-stemming.

The output accuracy is assured if the output word is present as root word in the database. The accuracy was calculated by dividing the number of correct outputs by the total number of inputs. The calculated average output accuracy of the stemmer was 81.27%.

4.3. Stemmer for Only Nouns & Proper Names

Gupta and Lehal [12] proposed a technique to stem nouns. Proper nouns are the words that are unavailable in dictionaries. A total of 17598 words were identified as proper names from Punjabi corpus which covered 13.84% of the entire corpus.

The original stemming algorithm composed of 19 steps and in first 18 steps suffix removal and/or suffix substitution was done based on the suffix attached to the word. Some suffixes were just removed from the end of the word but in some cases, suffix was removed and then another suffix was substituted. In 19th step of the algorithm, the stemmed word was searched in the list. If matched, it was taken as a Punjabi noun or proper name. The Algorithm 3 shows the reconstructed steps taken by the stemming process. Figure 3 shows these steps graphically.

Algorithm 3: Reconstructed Algorithm of Stemmer for Only Nouns & Proper Names
1: Input Punjabi word for Stemming.
2: Match end of the word with expected suffix.
3: If Not Matched; More expected suffix exist; goto (2) with next expected suffix.
4: Remove matched suffix from word.
5: If required, substitute a suffix.
6: Find word in noun morph/names list.
7: If Not Found; goto (9).
8: Display stem word.
9: End.

The stemmer was analyzed with 50 Punjabi documents obtained from various Punjabi news articles and other resources. It was tested to have 87.37% accuracy rate. The accuracy was calculated by dividing the number of correct outputs by the total number of inputs. The error percentage of the stemmer was also analyzed. The error percentage was also calculated separately for three different reasons.
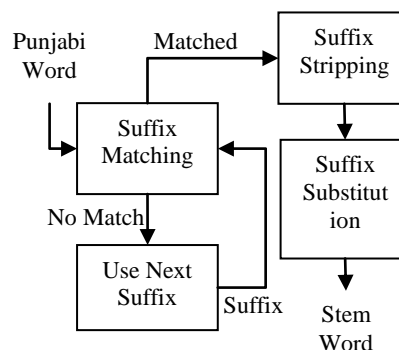

Figure 3: Stemmer for Only Nouns & Proper Names

4.4. Automatic Stemmer for All Category of Words

Gupta [13] proposed a complete stemmer for each category of Punjabi language words including proper names, nouns, adjectives, verbs, pronouns and adverbs. The suffixes in each category are identified manually by analysis of Punjabi news corpus.

For each category of suffixes separate rule based algorithm was written. But for stemming of proper names and nouns, the researcher reused the stemming algorithm from his already published paper [12]. The Punjabi word was taken by the algorithm and was searched in Punjabi dictionary. If matched, the same word was returned as stem word. Otherwise, four separate sub-stemmers were used one by one starting with the "Proper Names and Nouns Stemmer". The working of other three sub-stemmers was similar to that stemmer except that they work for separate category of words. After each sub-stemmer was applied, the output word was searched in a dictionary. If the word was matched in dictionary then it was taken as the output stem word, otherwise an error message was displayed due to the inability to find the actual stem word. The overall stemming process takes the steps shown in Algorithm 4. Figure 4 represents the steps graphically.

Algorithm 4: Automatic Stemmer for All Category of Words

1: Input Punjabi Word for stemming.
2: Find Word in noun morph OR Names dictionary OR Dictionary.
3: If Found; goto (17).
4: Call (Proper Names and Nouns Sub-Stemmer).
5: Find Stemmed Word in noun morph OR Names dictionary.
6: If Found; goto (17).
7: Call (Verb Sub-Stemmer).
8: Find Stemmed Word in dictionary.
9: If Found; goto (17).
10: Call (Adjective/Adverb Sub-Stemmer)
11: Find Stemmed Word in dictionary.
12: If Found; goto (17).
13: Call (Pronoun Sub-Stemmer)
14: Find Stemmed Word in dictionary.
15: If Found; goto (17).
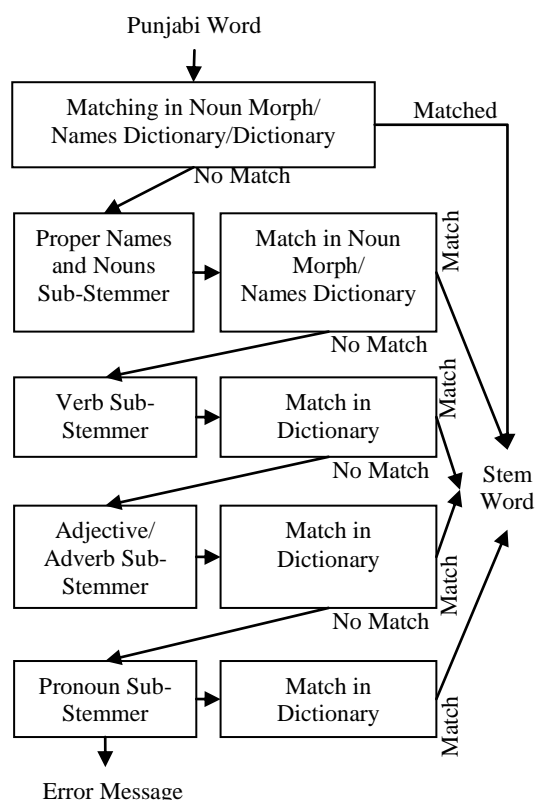*16: Show Error.*
17: Display Word.
18: End.

Figure 4: Automatic Stemmer for All Category of Words

The stemmer's accuracy percentage as well as error percentage was recorded for each category of sub-stemmer algorithm separately.

The overall accuracy percentage of the stemmer was 87.2% and error percentage was 12.8%. The accuracy was calculated by dividing the number of correct outputs by the total number of inputs.

4.5. Stemmer using WordNet Database

Puri et al. [14] proposed a modified version of Punjabi stemmer that was developed by Gupta et al. [12]. They utilized the Punjabi Wordnet database. A Punjabi dictionary database was prepared from Punjabi Wordnet taken from Indian Language Technology, IIT Bombay and Punjabi news articles. The Wordnet provided 53902 unique words of Punjabi language and 147942 distinct Punjabi words were obtained from news articles. The purpose of Punjabi dictionary database was to match the stemmed word to ensure the correctness of stemming.

A Named Entity database was prepared with 26992 distinct names used in Punjabi language. Most of the words were taken from news articles and university rolls. This list was prepared to ensure that these words should be skipped while stemming, because stemming of these Named Entities will change them altogether. A suffix list consisting of 75 Punjabi language suffixes was created manually by analyzing news articles. These are the suffixes to be removed and the suffixes to be substituted (if required) for stemming. The substitution suffix was NULL if there was no need of any suffix substitution.

Some of the suffix stripping rules provided false positive in stemming for a few words, yet those rules were important to stem a good quantity of words. Those rules were analyzed and an Exception list of those words was generated.

The algorithm proceeds by checking the word length. If its length is less than the minimum length (constant) then it need not be stemmed. If it is not, the word is searched in Named Entity database. If a match occurs, the word need not be stemmed. If the word is not found in Named Entity database, then it is searched in Exceptions list. If a match occurs, the word is ignored and not stemmed. If the word is not found in Exceptions list then its suffix is removed by searching the suffix in the suffix list. The substitution suffix (if available

in suffix list) is also added to the end of the word. Finally the stemmed word is searched in dictionary. The word is taken as stem word if it is available. Algorithm 5 shows the stemming process and Figure 5 graphically represents the process.

Algorithm 5: Stemmer using WordNet Database
1: Input Punjabi Word for stemming.
2: If Length (Word) < Min-Length; goto(9)
3: If Word Found in Named Entity List; goto(9)
4: If Word Found in Exception List; goto(9)
5: If Word Suffix Not Exist in Suffix List; goto(9)
6: Replace Suffix with Substitution Suffix
7: If Stemmed Word Not exist in dictionary; goto(9)
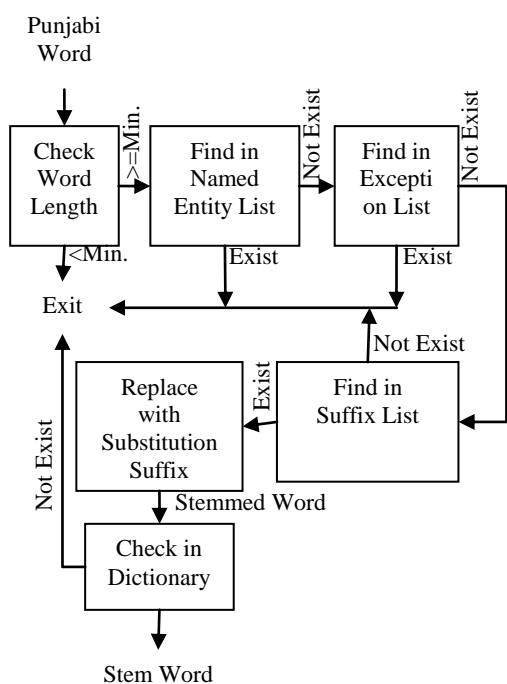*8: Output Stem Word*
9: End.

Figure 5: Stemmer using WordNet Database

The stemmer was tested with a sample of 1000 news articles. It was tested by computing the F-Measure value using the counting of words as follows:

$$C = correctly\ stemmed$$

$$U = under\text{-}stemmed$$

$$O = over\text{-}stemmed$$

$$Recall = C / (C + U) \qquad (1)$$

$$Precision = C / (C + O) \qquad (2)$$

$$The\ calculated\ Recall = 94.83\%$$

$$Precision = 90.84\%$$

From (1) and (2):

$$F\text{-}Measure = 92.79\%.$$

5. Discussion and Analysis

From the analysis of these stemmers, it is clear that each stemmer developed for Punjabi language provides accuracy improvements in its own way. The stemmer by Puri et al. [14] claims the most accurate stemmer according to the F-Measure calculations done by these researchers. But it is clear that [14] did not provide accuracy percentage as done by other researchers. By other researchers, the accuracy was calculated by dividing the number of correct outputs by the total number of inputs.

We can calculate the accuracy of stemmer developed by Puri et al. [14] using this formula for better comparison. Let the counting of words as follows:

$$C = correctly\ stemmed$$

$$U = under\text{-}stemmed$$

$$O = over\text{-}stemmed$$

$$So,\ Input\ Word\ Count = C + U + O$$

Using Calculations by Puri et al. [14], from equations (1) and (2) we have:-

$$Recall = C / (C + U) \times 100 = 94.83$$

$$i.e.\ C = 0.9483C + 0.9483U$$

$$0.0517C = 0.9483U$$

$$U = 0.0545186123C \qquad (3)$$

$$Precision = C / (C + O) \times 100 = 90.84$$

$$i.e.\ C = 0.9084C + 0.9084O$$

$$0.0916C = 0.09084O$$

$$O = 0.100836636C \qquad (4)$$

Using the values from equations (3) and (4), we have:-

$$Accuracy = C / (C + U + O) \times 100$$

$$= C / (C + 0.0545186123C + 0.100836636C) \times 100$$

$$= C / (1.155355248C) \times 100$$

$$= 86.553465 = 86.55 \text{ (approx.)}$$

The graph in Figure 6 shows the accuracy percentage of each Punjabi language stemmer calculated with same formula.
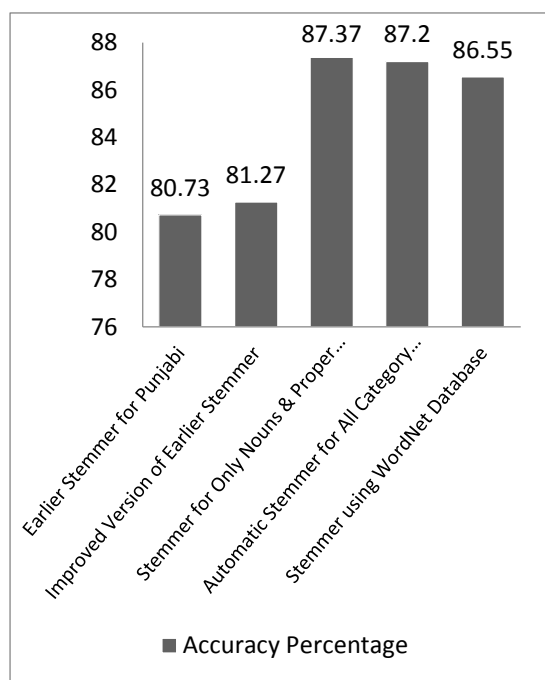

Figure 6: Graph Showing Accuracies of Punjabi Stemmers

From the graph it is clear that the overall accuracy of stemmers is reduced continuously after the stemmer [12]. But the stemmer [12] has its own limitations. It deals with only nouns and proper names and does not stem other category of words. So its accuracy is not comparable to those stemmers that cover all categories of words. This same stemmer is reused by Gupta [13] to develop a stemmer for all categories of words. But if we read original paper by Gupta [13], it discusses more and more about the "stemmer for nouns and proper names" and does not explore new research to that extent. In the abstract, the researcher mentioned the accuracy of only reused stemmer for nouns and proper names. Although the algorithms for verb, adjective/adverb and pronoun are explained in separate sections but those are not mentioned in results and discussion section except from the accuracy figures shown in a table. The whole 'results and discussion' section details about the same reused stemmer. The conclusion section of the paper also concludes with the reference of "stemmer for nouns and proper names". It means this field of research demands for more efficient approaches that could cope with limitations of existing stemmers.

6. Conclusion

Stemming is a building block of NLP. Research to stem Punjabi words is limited to a handful of research papers. The initiatives by researchers, no doubt, are praiseworthy, but this research field requires more efforts as compared to the research for other languages such as English, Hindi etc. Each stemmer developed for Punjabi language provides accuracy improvements in its own way. The table lookup approach is used in almost all Punjabi language stemmers to check either the word is already a stem word or the word is correctly stemmed after stemming process. For these stemmers, table lookup is the only way to check for the correctness of the stemming process. The rule based approach is also used along with table lookup to obtain a stem word by suffix stripping from the inflected words. Suffix substitution is used to complete the word, if required. This paper explored these approaches in an analytical manner with their accuracy comparisons. The significance of this analysis is that it disclosed the gap in NLP research for Punjabi language. There is a scope of research to invent more efficient stemmers for Punjabi language using new ideas or by extending the existing approaches.

7. References

[1] Martin Braschler, Barbel Ripplinger, "How Effective is Stemming and Decompounding for German Text Retrieval?", Information Retrieval, 7, 2004, pp. 291–316.

[2] Majumder P., Mitra M., Datta K., 'Statistical vs. Rule-Based Stemming for Monolingual French Retrieval", In: Peters C. et al. (eds) Evaluation of Multilingual and Multi-modal Information Retrieval. CLEF 2006. Lecture Notes in Computer Science, vol 4730. Springer, Berlin,

Heidelberg, 2007. doi:10.1007/978-3-540-74999-8_14

[3] Leena Jain, Prateek Agrawal, 'Text independent root word identification in Hindi language using natural language processing", International Journal of Advanced Intelligence Paradigms (IJAIP), Vol. 7, No. 3/4, 2015. doi: 10.1504/IJAIP.2015.073705

[4] Chandra Shakhar Yadav, Aditi Sharan, "Automatic Text Document Summarization Using Graph Based Centrality Measures on Lexical Network", International Journal of Information Retrieval Research, 8, 3 (July 2018), 14–32. doi: 10.4018/IJIRR.2018070102

[5] Daniel Gros, Christine Meschede, Tim Habermann, Giulia Kirstein, S. Denise Ruhrberg, Adrian Schmidt, and Tobias Siebenlist, "Anaphora Resolution: Analysing the Impact on Mean Average Precision and Detecting Limitations of Automated Approaches", International Journal of Information Retrieval Research 8, 3 (July 2018), 33–45. doi: 10.4018/IJIRR.2018070103

[6] Prateek Agrawal, Leena Jain, "Anuvaadika: Implementation of Sanskrit to Hindi Translation Tool Using Rule-Based Approach", Recent Advances in Computer Science and Communications (2020) 13: 1136. doi: 10.2174/2213275912666181226155829

[7] Bhadwal N., Agrawal P., Madaan V., "Bilingual Machine Translation System Between Hindi and Sanskrit Languages", In: Luhach A., Jat D., Hawari K., Gao XZ., Lingras P. (eds) Advanced Informatics for Computing Research. ICAICR 2019. Communications in Computer and Information Science, vol 1075. Springer, Singapore, 2019. doi: 10.1007/978-981-15-0108-1_29

[8] Bhadwal N., Agrawal P., Madaan V., "A Machine Translation System from Hindi to Sanskrit Language using Rule based Approach", Scalable Computing: Practice and Experience, Vol. 21, pp. 543-554.doi: 10.12694/scpe.v21i3.1783

[9] Abdullah Wahbeh, Mohammed Al-Kabi, Qasem Al-Radaideh, Emad Al-Shawakfa, Izzat Alsmadi, "The Effect of Stemming on Arabic Text Classification: An Empirical Study" International Journal of Information Retrieval Research, Vol. 1, No. 3, 2011. doi: 10.4018/IJIRR.2011070104.

[10] Dinesh Kumar, Prince Rana, "Design and Development of a Stemmer for Punjabi", International Journal of Computer Applications (ISSN: 0975 – 8887), Vol.11, No.12, December 2010, pp.18-23.

[11] Dinesh Kumar, Prince Rana, "Stemming of Punjabi Words by using Brute Force Technique", International Journal of Engineering Science and Technology (IJEST) ISSN: 0975-5462, Vol. 3 No. 2, Feb 2011, pp. 1351-1358.

[12] Vishal Gupta, Gurpreet Singh Lehal, "Punjabi Language Stemmer for nouns and proper names", Proceedings of the 2nd Workshop on South and Southeast Asian Natural Language Processing (WSSANLP), IJCNLP 2011, Chiang Mai, Thailand, November 8, 2011, pp.35–39.

[13] Vishal Gupta, "Automatic Stemming of Words for Punjabi Language", Advances in Signal Processing and Intelligent Recognition Systems-Vol. 264, Springer International Publishing Switzerland, 2014, pp.73-84, doi:10.1007/978-3-319-04960-1_7

[14] Rajeev Puri, R. P. S. Bedi, Vishal Goyal, "Punjabi Stemmer Using Punjabi WordNet Database", Indian Journal of Science and Technology, Vol. 8(27), October 2015, pp.1-5, doi: 10.17485/ijst/2015/v8i27/82943.