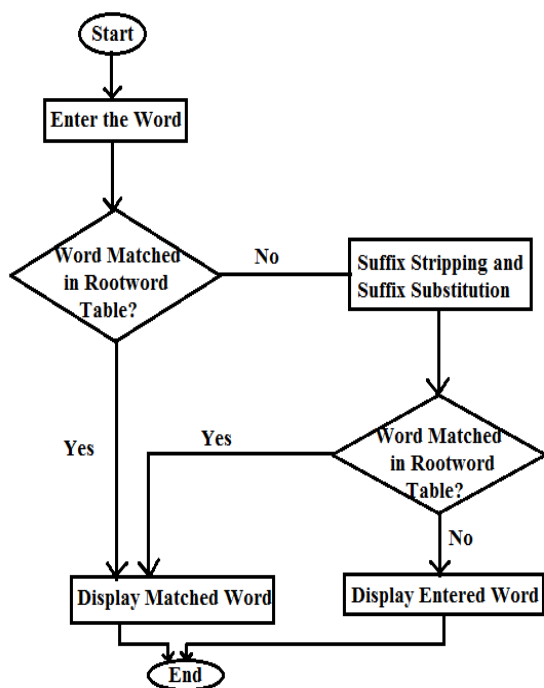


entered word is root word or not. If it is found then it will be simply displayed as output, which means the word entered by the user is root word. If it doesn't exist in root word database then suffix stripping along with suffix substitution will be applied.



In the proposed stemmer there will be two database table one table will be of root words having approximately 250000 root words and the other table will maintain the suffix to be stripped and corresponding suffixes to be substituted. So suffix stripping and suffix substitution will be done from the database of suffixes. After that the stemmed word will again be checked in the root word database table to ensure that the word is stripped correctly and will be displayed to the user. The proposed stemmer reduces the problem of over-stemming and under-stemming since the result is displayed to user after ensuring the stemmed word in the root word database i.e. if the stemmed word is over-stemmed or under-stemmed it will not match in the root word database. So, accurate results will be displayed to user.

6. Conclusion and Future Work

The proposed stemmer will overcome the problem of over-stemming and under-stemming by using suffix substitution and suffix stripping algorithm. Further a new hybrid approach can be proposed by combing some other algorithms.

7. References

- [1] Dinesh Kumar and Prince Rana (2010) "Design and Development of Stemmer for Punjabi", International Journal of Computer Applications (0975 – 8887)
- [2] Iliia Smirnov (2008) "Overview of Stemming Algorithms"
- [3] Mohd. Shahid Husain (2012), "An unsupervised approach to develop stemmer", International Journal on Natural Language Computing (IJNLC)
- [4] Prince Rana and Dinesh Kumar (2011) "stemming of Punjabi words by using brute force technique", International Journal of Engineering Science and Technology (IJEST) ISSN
- [5] Ramanathan, Ananthkrishnan, and Durgesh D. Rao. (2003). "A Lightweight Stemmer for Hindi", Workshop on Computational Linguistics for South-Asian Languages, EACL.
- [6] Upendra Mishra and Chandra Prakash (2012) "MAULIK: An effective stemmer for Hindi language" International Journal on Computer Science and Engineering (IJCSE)