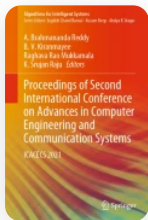


[Home](#) > [Proceedings of Second International Conference on Advances in Computer Engineering and Communication Systems](#) > Conference paper

Design and Development of Marathi Word Stemmer

| Conference paper | First Online: 22 February 2022

| pp 35–48 | [Cite this conference paper](#)



Proceedings of Second International Conference on Advances in Computer Engineeri...

P. Vaishali Kadam, B. Kalpana Khandale & C. Namrata Mahender

 Part of the book series: [Algorithms for Intelligent Systems \(\(AIS\)\)](#)

 559 Accesses

Abstract

Stemming is a basic morphological analysis tool or a process used to remove variants of a particular word and to identify its root or stem. In this, morphological analysis is carried out to identify various forms of a word to obtain its stem. Stemming is used as a preprocessing operational tool for the development of various natural language text applications, such as part-of-speech tagging, sentiment analysis, text segmentation, text classification, text summarization, information extraction, information retrieval

applications, and named entity recognition. Stemming is generally used to achieve better performance and evaluation of information retrieval systems. It is achieved by increasing the standard in the format of the text. There may be various forms of the same root word or base that are used in the system database to increase storage requirements by stemming. Index files are reduced to the words present in the text with their linguistic variants in the documents. It is useful for memory management and system performance and for reducing memory requirements. Our proposed system is designed specifically for Marathi language text because for some other languages in Deonagri scripts like Hindi or Konkani stemmers are available and have good results and performance. But for Marathi, very little work has been done on stemmers. The proposed idea is to develop a stemmer for the Marathi language by using a supervised machine learning approach and our own developed handwritten grammar rule set and word dictionary using which the suffixes of the different variants are successfully removed for generating their respective stem words. Finally, the system is evaluated for its performance. It is measured at 61.34%.

 This is a preview of subscription content, [log in via an institution](#)  to check access.

Access this chapter

[Log in via an institution](#)

Chapter

EUR 29.95

Price includes VAT (India)

Available as PDF

Read on any device

Instant download

Own it forever

[Buy Chapter](#) 

eBook

EUR 192.59

▼ **Softcover Book**

EUR 229.99

▼ **Hardcover Book**

EUR 229.99

Tax calculation will be finalised at checkout

Purchases are for personal use only

[Institutional subscriptions](#) →

References

1. H.B. Patil, B.V. Pawar, A.S. Patil, A comprehensive analysis of stemmers available for Indic languages. *Int. J. Nat. Lang. Comput. (IJNLC)* 5 (2016)
[Google Scholar](#)
2. Md. Redowan Mahmud, M. Afrin, Md. Abdur Razzaque, E. Miller, J. Iwashige, A rule based Bengali stemmer, in *International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (IEEE, 2014). 978-1-4799-3080-7/14
[Google Scholar](#)
3. I.R. Vijaya Lakshmi, S. Britto Ramesh Kumar, Literature review: stemming algorithms for Indian and non-Indian languages. *Int. J. Adv. Res. Comput. Sci. Technol. (IJARCST)* 2(3) (2014)
[Google Scholar](#)
4. U. Mishra, C. Prakash, MAULIK: an effective stemmer for Hindi language. *Int. J. Comput. Sci. Eng. (IJCSE)* (2012)

5. <https://www.researchgate.net/publication/233955548>

6. D. Bijal, S. Sanket, Overview of stemming algorithms for Indian and non-Indian languages. (IJCSIT) Int. J. Comput. Sci. Inf. Technol. 5(2), 1144–1146 (2014)

7. A. Ramanathan, D.D. Rao, *A Lightweight Stemmer for Hindi* (KBCS CDAC Mumbai, 2019). <http://www.kbcs.in/downloads/papers/StmmerHindi.pdf>

8. Md. Shahidul Salim (Shakib), T. Ahmed, K.M. Azharul Hasan, A formal method for designing a Bangla Stemmer using rule based approach. Conference paper (2019). <https://www.researchgate.net/publication/337703586>

9. S. Sarkar, S. Bandyopadhyay, Morpheme extraction task using Mulaadhaar—a rule-based stemmer for Bengali (2012). isical.ac.in/~fire/data/working-notes_2012/met/MET_JU.pdf

10. D. Ganguly, J. Leveling, G.J.F. Jones, Rule-based stemmers for Bengali and Hindi, in *FIRE 2012 Workshop*, Kolkata, India (2012)

11. T.T. Urmi, J.J. Jammy, S. Ismail, *A corpus based unsupervised Bangla word stemming using N-Gram language model* (2016). <https://www.researchgate.net/publication/312430235>

12. J. Sheth, B. Patel, A stemmer for morphological level analysis of Gujarati language, in *International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, Ghaziabad, India (2014)

[Google Scholar](#)

13. A.K. Barman, J. Sarmah, S.K. Sarma, Development of Assamese rule based stemmer using WordNet, in *Proceedings of the 10th WordNet Conference*, Wroclaw, Poland (2019)

[Google Scholar](#)

14. J. Ameta, N. Joshi, I. Mathur, A lightweight stemmer for Gujarati (2012).
ResearchGate publication/232416333

[Google Scholar](#)

15. D. Kumar, P. Rana, Stemming of Punjabi words by using brute force technique. *Int. J. Eng. Sci. Technol. (IJEST)* 3 (2011).

<https://www.researchgate.net/publication/50406900>

16. V. Gupta, G.S. Lehal, Punjabi language stemmer for nouns and proper names, in *Proceedings of the 2nd Workshop on South and Southeast Asian Natural Language Processing (WSSANLP), IJCNLP*, Chiang Mai, Thailand (2011), pp. 35–39

[Google Scholar](#)

17. P. Kaur, P.K. Buttar, A rule-based stemmer for Punjabi verbs. *Int. Res. J. Eng. Technol. (IRJET)* 06(05) (2019)

[Google Scholar](#)

18. N. Pise, V. Gupta, Rule based stemmer for Marathi language. *Int. J. Comput. Sci. Eng.* 6(5) (2018). E-ISSN: 2347-2693

19. P. Pandey, Int. J. Adv. Res. Comput. Commun. Eng. 5(10) (2016)

[Google Scholar](#)

20. M.M. Majgaonker, T.J. Siddiqui, Discovering suffixes: a case study for Marathi language. Int. J. Comput. Sci. Eng. (IJCSE) 02, 2716–2720 (2010)

[Google Scholar](#)

21. M.R. Valambe, *Sugam Marathi Vyakaran va Lekhan*, 54th edn. (Nitin Publication, 2019)

[Google Scholar](#)

Acknowledgements

Authors would like to acknowledge and thank CSRI DST Major Project sanctioned No. SR/CSRI/71/2015 (G), Computational and Psycholinguistic Research Lab Facility supporting this work and Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India and also thankful to Chhatrapati Shahu Maharaj Research Training and Human Development Institute (SARTHI), Pune, for providing financial assistance for this Ph.D. research work. I would like to express my sincere thanks to research guide Dr. C. Namrata Mahender (Asst. Professor) of the Department of Computer Science and IT, Dr. B. A. M. U., Aurangabad, for providing research facilities, constant technical, and moral support.

Author information

Authors and Affiliations

Department of Computer Science & I.T., Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India

P. Vaishali Kadam, B. Kalpana Khandale & C. Namrata Mahender

Editor information

Editors and Affiliations

Department of Computer Science and Engineering, Vallurupalli Nageswara Rao
Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India
A. Brahmananda Reddy

Department of Computer Science and Engineering, Vallurupalli Nageswara Rao
Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India
B.V. Kiranmayee

Centre for Business Data Analytics, Copenhagen Business School, Frederiksberg,
Denmark
Raghava Rao Mukkamala

Department of Computer Science and Engineering, CMR Technical Campus,
Hyderabad, Telangana, India
K. Srujan Raju

Rights and permissions

[Reprints and permissions](#)

Copyright information

© 2022 The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd.

About this paper

Cite this paper

Vaishali Kadam, P., Kalpana Khandale, B., Namrata Mahender, C. (2022). Design and Development of Marathi Word Stemmer. In: Reddy, A.B., Kiranmayee, B., Mukkamala, R.R., Srujan Raju, K. (eds) Proceedings of Second International Conference on Advances in Computer Engineering and Communication Systems. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-16-7389-4_4

[.RIS](#)  [.ENW](#)  [.BIB](#) 

DOI

https://doi.org/10.1007/978-981-16-7389-4_4

Published

22 February 2022

Publisher Name

Springer, Singapore

Print ISBN

978-981-16-7388-7

Online ISBN

978-981-16-7389-4

eBook Packages

Intelligent Technologies
and Robotics

Intelligent Technologies
and Robotics (R0)

Publish with us

[Policies and ethics](#) 