



GUJARAT TECHNOLOGICAL UNIVERSITY

(Established Under Gujarat Act No.: 20 of 2007)

ગુજરાત ટેકનોલોજીકલ યુનિવર્સિટી

(ગુજરાત અધિનિયમ ક્રમાંક : ૨૦/૨૦૦૭ દ્વારા સ્થાપિત)

Abstract of the Thesis



Name: MODH JATINKUMAR CHAMANLAL

Enrollment No: 169999902001

Faculty: Computer Science

Discipline/Branch: Computer Science

Title of the Thesis: A System for Simplification of Idiomatic Gujarati Text for Improved Interlingual Language Processing

Abstract

All existing Gujarati machine translation systems, including Microsoft Translator and Google Translate face the problem with the idiomatic Gujarati text. They are unable to properly translate Gujarati idioms. The proposed system simplifies the Gujarati idioms by correctly recognizing all Gujarati idiom phrases present in the input and replacing them with the corresponding Gujarati meanings of the idiom. As part of interlingual processing, this model translates all Gujarati idioms into the same Gujarati language but with simplified Gujarati meaning. The result provided by the proposed system is simplified Gujarati text that does not contain any Gujarati idioms.

Overall 3472 distinct and 6081 non-distinct Gujarati idioms are collected, analyzed and classified. This research work classifies the Gujarati idioms into N-gram, M-meaning, root idioms, inflected idioms and personage idioms. Because Gujarati idioms are used in a variety of formats and in a variety of contexts in real life, recognizing them all can be a difficult job for any machine translation system. The proposed system detects all inflected and static idiom formats from the Gujarati text by employing a dictionary-based approach and a rule-based approach to generate dynamic idiom forms. Dynamic generation and detection of idioms are possible using the newly generated 15 suffix and diacritics-based rules. In the case of multiple meaning idioms, a context-based search algorithm determines the particular meaning of the idiom using surrounding contextual words. In addition, the readability complexity prediction model calculates the readability complexity score and predicts the complexity type for the idiomatic Gujarati text by considering four different parameters. This is innovative and the first of its kind in the Gujarati language research community.

This PhD Thesis will help in the translation of Gujarati idioms into any other language in the world. This research contributes to the simplification of idiomatic Gujarati text by detecting and replacing all Gujarati idioms with equivalent Gujarati meanings. This research work helps in dealing with Gujarati idiomatic text and also calculates the readability complexity score for it.

List of Publication(s):

- 1) M.K. Audichya, J.R. Saini and J.C. Modh, "Towards a Richer IndoWordNet with New Additions for Hindi and Gujarati Languages", International Journal of Advanced Computer Science and Applications(IJACSA), 13(7), 2022, pg837-842, <http://dx.doi.org/10.14569/IJACSA.2022.0130797>
- 2) J.C. Modh, J.R. Saini and K. Kotecha, "A Novel Readability Complexity Score for Gujarati Idiomatic Text", International Journal of Advanced Computer Science and Applications(IJACSA), 13(5), 2022, pg453-459, <http://dx.doi.org/10.14569/IJACSA.2022.0130553>
- 3) J.C. Modh, J.R. Saini and K. Kotecha, "A Novel Morphological Analysis based Approach for Dynamic Detection of Inflected Gujarati Idioms", International Journal of Advanced Computer Science and Applications(IJACSA), 13(4), 2022, pg186-193, <http://dx.doi.org/10.14569/IJACSA.2022.0130422>
- 4) J.C. Modh and J.R. Saini, "A Novel Hybrid Translator for Gujarati to Interlingual English MTS for Personage Idioms", Smart Trends in Computing and Communications, Proceedings of SmartCom 2022. Lecture Notes in Networks and Systems, vol 396. Springer, Singapore. pg355-367, https://doi.org/10.1007/978-981-16-9967-2_34
- 5) J.C. Modh and J.R. Saini, "Dynamic Phrase Generation for Detection of Idioms of Gujarati Language using Diacritics and Suffix-based Rules", International Journal of Advanced Computer Science and Applications(IJACSA), 12(7), 2021, pg239-247, <http://dx.doi.org/10.14569/IJACSA.2021.0120728>
- 6) J.C. Modh and J.R. Saini, "Using IndoWordNet for Contextually Improved Machine Translation of Gujarati Idioms", International Journal of Advanced Computer Science and Applications(IJACSA), 12(1), 2021, pg225-232, <http://dx.doi.org/10.14569/IJACSA.2021.0120128>
- 7) J.C. Modh and J.R. Saini, "Context Based MTS for Translating Gujarati Trigram and Bigram Idioms to English", 2020 International Conference for Emerging Technology (INCET), 2020 IEEE, pp. 1-6, doi: 10.1109/INCET49848.2020.9154112
- 8) J.C. Modh and J.R. Saini, "A Study of Machine Translation Approaches For Gujarati Language", International Journal of Advanced Research in Computer Science, (IJARCS) UGC Approved, Peer Reviewed Journal, Volume 9 Issue 1, 285-288, 2018. doi:<https://doi.org/10.26483/ijarcs.v9i1.5266>
- 9) J.C. Modh and J.R. Saini, "Study of Machine Translation Systems and Techniques for Indian Languages", Proceeding of International Conference The Journey of Indian Languages: Perspectives on Culture and Society Vol - 2 (English), 2019. pg22-30, <http://gyansampada.baou.edu.in:8080/jspui/handle/123456789/458>
- 10) J.R. Saini and J.C. Modh, "GIdTra: A dictionary-based MTS for translating Gujarati bigram idioms to English," 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2016, pp. 192-196, doi: 10.1109/PDGC.2016.7913143