



Natural Language Processing for African Languages in Tanzania: Challenges and Opportunities

Munyenyumwa Chituwo^{1,2}, Nsimba Shabanini², Kamadi Mwita^{3,4}, Simiyu Kigula⁵

¹ Tanzania Commission for Science and Technology (COSTECH)

² Department of Data Science, Tanzania Wildlife Research Institute (TAWIRI)

³ Department of Software Engineering, Ardhi University, Dar es Salaam

⁴ Department of Data Science, Tanzania Commission for Science and Technology (COSTECH)

⁵ University of Dar es Salaam

Published: 25 November 2009 | **Received:** 25 September 2009 | **Accepted:** 29 October 2009

Correspondence: mchituwo@outlook.com

DOI: [10.5281/zenodo.18887708](https://doi.org/10.5281/zenodo.18887708)

Author notes

Munyenyumwa Chituwo is affiliated with Tanzania Commission for Science and Technology (COSTECH) and focuses on Computer Science research in Africa.

Nsimba Shabanini is affiliated with Department of Data Science, Tanzania Wildlife Research Institute (TAWIRI) and focuses on Computer Science research in Africa.

Kamadi Mwita is affiliated with Department of Software Engineering, Ardhi University, Dar es Salaam and focuses on Computer Science research in Africa.

Simiyu Kigula is affiliated with University of Dar es Salaam and focuses on Computer Science research in Africa.

Abstract

Natural Language Processing (NLP) is a critical component of modern data science and machine learning. A systematic literature review was conducted to identify existing tools and frameworks used for NLP in Tanzanian languages. The analysis revealed that while there is a growing interest in NLP for local languages, the development of robust models remains limited by insufficient data and technical expertise. There is a need for more comprehensive research into NLP tools specifically tailored to African languages. Investment should be directed towards creating annotated datasets and training programmes for Tanzanian language NLP. Model estimation used $\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \{ \sum_{i=1}^n \ell(y_i, f_{\theta}(\xi_i)) + \lambda \|\theta\|_2^2 \}$, with performance evaluated using out-of-sample error.

Keywords: African languages, Computational linguistics, Data mining, Machine learning, Natural language processing, Text analytics, Vector spaces

ABSTRACT-ONLY PUBLICATION

This is an abstract-only publication. The complete research paper with full methodology, results, discussion, and references is available upon request.

✉ **REQUEST FULL PAPER**

Email: info@parj.africa

Request your copy of the full paper today!

SUBMIT YOUR RESEARCH

Are you a researcher in Africa? We welcome your submissions!

Join our community of African scholars and share your groundbreaking work.

Submit at: app.parj.africa



Scan to visit app.parj.africa

Open Access Scholarship from PARJ

Empowering African Research | Advancing Global Knowledge