PAPER REVIEW FORM

Paper ID: 1

Paper Title: Shorthand Romanized Tamil to Tamil Transliteration Using Novel Hybrid Approach

Reviewer's Index Number: 20000499

(Please mark with X the most appropriate response)

	Yes	No
1. Relevance		
1.1. This paper is relevant, timely, and of interest to the audience who are interested to see the development in Computing	X	

	Yes	No
2. Novelty or Originality		
2.1. The paper presents new, innovative or insightful information	X	
2.2. The paper reflects current information on this topic	X	

	Yes	No
3. Title		
3.1. The title is clear and informative	X	
3.2. The title reflects the content and purpose of the paper	X	

	Yes	No
4. Abstract		
4.1. The abstract is concise and relevant	X	

	Yes	No
5. Keywords		
5.1. The keywords provide adequate index entry for the paper (up to 4 words)		X

6. Content

	Yes	No
6.1. Introduction		
6.1.1. The purpose and objectives are clearly stated in the introductory section		X

If 'no', please give reasons

The purpose is mentioned there but it does not clearly state the objectives. Introduction section should explain the problem properly and highlight the gap we are going to fill. These objectives are not clearly mentioned in the introduction.

	Yes	No
6.2. Background/Literature Study		
6.2.1. The most current references on this topic have been included	X	
6.2.2. The most relevant references on this topic have been included	X	

If 'no', please give reasons

6.2.1 – Most of the references are from 2020,2021 and 2022. Hence the most current references on this topic have been included.
6.2.2 - Most of the references provide details useful for the methodology hence the most relevant references on this topic have been included

	Yes	No
6.3. Methodology		
6.3.1. The research methodology for the study is appropriate and applied properly	⌗	

If 'no', please give reasons

Yes
This paper clearly mentioned the methodology using several sub-topics such as Data analysis, Data Annotation, Data collection methods, and the main approach used for the conversion of Romanized Tamil to Tamil. It used several diagrams and test results to clearly mention the idea. Also, it provides some statistical results that are used to check the accuracy of the model.

	Yes	No
6.4. Results		
6.4.1. Results are methodologically and correctly presented	X	
6.4.2. Results are relevant to the problem addressed	X	
6.4.3. Discussion of the results is based on analysis of data		X
6.4.4. Results are not overstated or overgeneralized	X	

If 'no', please give reasons

6.4.1- The BLEU score and the word level accuracy have been used for evaluation purposes and it presents all the datasets and equations used for the evaluation with the test results. Hence the results are methodologically and correctly presented.
6.4.2 - Since the paper discussed the efficient method for translating Romanized Tamil to Tamil the accuracy results are relevant to the problem addressed here.
6.4.3 - Discussion of the results is based on the word level accuracy and the character level accuracy, not on the analysis of data.
6.4.4 – The paper only provides word-level accuracy and character-level accuracy for the new mechanism and compares it with the previous method. Hence the results are not overstated or overgeneralized.

	Yes	No
6.5. Discussions and Conclusions		
6.5.1. Discussions and conclusions are presented in an appropriate manner	X	
6.5.2. Interpretation and conclusions are justified by the data		X
6.5.3. Implications and recommendations are relevant and useful	X	

If 'no', please give reasons

6.5.1 – In the conclusion the paper summarizes all the content and answers the research questions mentioned in the introduction part with the evidence results.
6.5.2 - The paper says that the TamZhi transliterator performs better than the existing rule-based method but it only provides the accuracy results for the TamZhi transliterator. So, we don't have enough statistical information to say that Tamzhi is better than the previous method.
6.5.2 – According to the achieved results the paper provides useful recommendations such as increasing the word level accuracy and accepting code-mixed Romanized Tamil etc.

	Yes	No
7. Clarity		
7.1. There are no any vague areas or parts which are difficult to understand		X
7.2. There are no any contradictions or inconsistencies	X	
7.3. The paper stays focused	X	

If 'no', please give reasons

> 7.1 – There are some vague parts such as when talking about the N-gram model. It uses the "UnigramTagger, BigramTagger, and TrigramTagger with a backoff mechanism from NLTK" which is difficult to understand without a prior knowledge of it.
> 7.2 – The paper does not mention the same thing in two different places differently.
> 7.3 – The paper does not provide any other information which is not related to the research.

	Yes	No
8. Organization		
8.1. Ideas are developed and related in a logic sequence		X
8.2. Transitions between discussions are smooth and easy to follow	X	

If 'no', please give reasons

> 8.1 – The paper provides a logical explanation for most of the parts starting from the introduction to the conclusion but there are some new terms when we are reading the paper that require prior knowledge of machine learning and neural computing. (Terms such as UnigramTagger, BigramTagger, and TrigramTagger with a backoff mechanism from NLTK)
> 8.2 – The flow of the system is clear. It starts with the introduction and next step-by-step explains the method and finally discusses the results and ends with the conclusion. The paper used several diagrams and tables for a clearer explanation.

	Yes	No
9. ACCURACY		
9.1. The supporting evidence (literature referenced) is appropriately cited		X
9.2. Tables and figures are of clear and satisfactory quality	X	
9.3. There are no math or text errors in tables or figures	X	
9.4. Legends and titles of tables and figures are clearly given	X	
9.5. The paper is free from grammatical or spelling errors	X	

If 'no', please give reasons

> 9.1-When the paper provides references for the BLEU score the link for the reference is not provided and the access date is not there for some of the references.
> 9.2-Tables clearly display the related information.
> 9.3- In this paper mathematical parts are used only for the accuracy calculation and it is mentioned clearly and also no text errors in the tables.

10. **Quality of the Paper** (mark an **'X'** for your recommendation):

Excellent	
Good	X
Fair	
Poor	

11. **Overall Recommendation** (mark an **'X'** for your recommendation):

Accept without further modifications	
Accept subject to minor revision	X
Reconsider after major revision	
Reject	

12. **Comments and recommendations to the Author(s)** (Please provide a detailed review, including justifications for your scores. These comments will be sent to the author(s). Therefore, please do not make any specific statement about the acceptability of the paper in this part):

The paper discussed the Romanized Tamil to Tamil transliteration and in the abstract section it is well organized and the problem is well defined. It provides a summary of the literature review including what the problem is briefly discusses the results and also provides the importance of the findings to the real world. In the introduction section it starts with discussing the term transliteration which is used in most of the places in the paper. So, it is really useful to discuss that there but I think it does not need that much explanation here. Because it is explained using several examples and it uses several paragraphs as well. Also, I think it's better to highlight the gap that the research going to fill and the objectives of the research in the introduction section.

The methodology is well explained using a step-by-step approach and using several diagrams and figures. However, for some figures, I think it's better to add a small explanation about it which gives a clearer understanding to the readers. Also, for some new terms such as the N-gram model, Unigram Tagger, Bigram Tagger, and NLTK it's better to include a reference that explains these terms. It will help readers without a prior understanding of machine learning and neural computing. Finally, the conclusion summarizes all the content and gives some valuable future enhancements. Also, I think it will be more accurate if they provide the accuracy levels for the existing rule-based method. So, the readers can compare those results and get a clearer idea about the TamZhi Transliterator. Overall the paper provides a good review of the problem.

13. **Comments to the editors** (This part is confidential, and is only presented to the editors):

The paper provides a detailed description of the problem that we have to face in the transliteration of Romanized Tamil to Tamil which is the missing vowels problem. So, it provided a hybrid approach to solve the problem combining rule-based techniques with trie modelling. From the introduction to the conclusion the paper is well organized and the methodology is well explained. However, there are some points which I think need to change a little bit. In the methodology, there are some new terms that are not familiar to the readers such as UnigramTagger, BigramTagger, and NLTK. So, I think it's better to add references for these terms. In the results and discussion section, the paper mentioned the BLEU score and it also has a reference as well. However, in the reference section, it is not properly referenced. For some of the references the accessed date is not there. As the final result, they show that there is a 93% character level accuracy and 60% word level accuracy. There is a remarkable difference between these 2 values. So it is better to discuss the reasons for this gap in the results section. The conclusion section of the paper has provided valuable future enhancements that can improve the current results. Adding the challenges that researchers has to face implementing future enhancements may add an additional value to the paper. Overall the paper provides valueable offerings for Roamnized Tamil to Tamil transliteration.