**Conference**

Paper ID: 12345  
• Title: A Novel Rule-based Recursive Stemming Algorithm for Plagiarism Detection in Devanagari Scripts

• Author(s): Ayush Kumar Shah

• Reviewer’s Recommendations:

i. **Writing** (choose one)   
o Not readable  
o Major improvement needed  
☑ Minor improvement suggested  
o Well written  
ii. **Novelty** (choose one)   
☑ Original  
o Somewhat interesting  
o Borderline   
o Been there, done that  
iii. **Suitability** (choose one)   
☑ Very related  
o Limited interests  
o Not suited  
iv. **Reviewer’s Expertise** (choose one)   
o Expert  
o Knowledgeable  
☑ Passing interests  
o Not my cup of tea  
v. **Recommendation** (choose one)  
o Absolute reject   
o Reject if there is no space  
☑ Accept if there is space  
o Absolute accept  
• Reviewer’s detailed comments: strength, weakness, suggestion  
i. **Comments for the Authors:**

The work is very solid and impressive. But I think in order for this paper to be published, you still need some revision. For example, you should explain more detailly about how to do lemmatization, such as why there still exists a check for suffix and prefix after previous two check about prefix and suffix? And another question, when do you choose to recombine suffix, but not prefix, is there a combination problem here? Besides, the Jaccard similarity you talk about in the end just concerns the vocabulary info, not the frequency info, right? I think you should explain more about these questions for the reader to get a better understanding. And if you want to improve your result, I think you should try some deep learning method in NLP to dig out more semantic and contextual information to construct feature vector and further compute the similarity extent.

ii. **Comments for the Program Committee (will be kept confidential and NOT released to the authors)**

This work deals with Nepali scripts to solve plagiarism detection problem. In my view, this work is well organized and using a solid NLP technique, besides, it also invents original grammatical rules to do lemmatization, which is very impressive. But given the fact that this paper only relies on root word frequency and category information in each document to do similarity computation. I am afraid that the experiment result may not be optimistic enough to be accepted.