Review for conference

Paper ID: X
Title: A Novel Rule-based Recursive Stemming Algorithm for Plagiarism Detection in Devanagari Scripts
Author: Ayush Kumar Shah

i. **Writing** (choose one)

- Major improvement needed
- Minor improvement suggested
- **Well written**

ii. **Novelty** (choose one)

- **Original**
- Somewhat interesting
- Borderline
- Been there, done that

iii. **Suitability** (choose one)

- **Very related**
- Limited interests
- Not suited

iv. **Reviewer's Expertise** (choose one)

- Expert
- **Knowledgeable**
- Passing interests
- Not my cup of tea

v. **Recommendation** (choose one)

- Absolute reject
- Reject if there is no space
- **Accept if there is space**
- Absolute accept

Reviewer's detailed comments: strength, weakness, suggestion

i. **Comments for the Authors:**

This paper preprocesses Devanagari scripts hoping to be used for multiple NLP tasks including plagiarism detection that was tested in this work. Preprocessing a foreign language text is a great contribution; however, for plagiarism detection, there are many more factors to consider. Presence of the common base forms of the words in two articles may not guarantee plagiarism. Some other factors that need to be considered for plagiarism detection might include, context, sentence level similarity, and document level similarity.

Also, the statistical measures reported are from 20 pairs of manually annotated articles. These measures might not hold true on real plagiarized texts. This work would have been improved if the author had trained and tested the algorithms on larger set of real plagiarized and non-plagiarized text.

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

This paper introduces a novel algorithm for plagiarism detection in Devanagari scripts.

Review for journal

- **Goals and Contributions:**
  i. Do the authors clearly state the research goals of the work?

     Yes

  ii. Does the paper clearly indicate what the contributions are?

     Yes

  iii. Are the claimed contributions original and significant in terms of
     - Novel methodology?
     - **New applications?**

  iv. Does the paper describe the methods in sufficient detail for readers to replicate the work?

     Yes

- **Evaluation:**
  i. Do the authors carefully evaluate the approach?

     Yes

  ii. Does the paper include systematic experiments, a careful theoretical analysis, or give evidence of generality?

     Yes

- **Discussion:**
  i. Does the paper discuss relevant earlier works, noting similarities, differences and progress?

     Yes

  ii. Does it discuss the limitation of the approach as well as its advantages?

     Yes

  iii. Does it consider the implication of the work and outline direction for future work?

     Yes

- **Presentation:**
  i. Is the paper properly organized and well written?

     Yes

  ii. Is the paper grammatically correct and free of spelling errors?

     Yes

  iii. Does it use standard terminology?

     Yes

- **Detailed Comments:**

  This paper preprocesses Devanagari scripts hoping to be used for multiple NLP tasks including plagiarism detection that was tested in this work. Preprocessing a foreign language text is a great contribution; however, for plagiarism detection, there are many more factors to consider. Presence of the common base forms of the words in two articles may not guarantee plagiarism. Some other factors that need to be considered for plagiarism detection might include, context, sentence level similarity, and document level similarity.

  Also, the statistical measures reported are from 20 pairs of manually annotated articles. These measures might not hold true on real plagiarized texts. This work would have been improved if the author had trained and tested the algorithms on larger set of real plagiarized and non-plagiarized text.

- **Recommendation:**
  i. The paper could be published in its current form.
  ii. The paper could be published after minor revision:
    - Another round of review is needed.
    - **No review is needed.**
  iii. The paper requires major revision for further consideration.
  iv. The paper is not suitable for publication in this journal.