International Journal of Electrical and Computer Engineering (IJECE)

#10995 Summary

Submission

Authors	Arbi Haza Nasution, Yohei Murakami, Toru Ishida
Title	Generating similarity cluster of Indonesian languages with semi-supervised clustering
Original file	10995-14652-2-SM.TEX 2018-01-11
Supp. files	10995-14656-1-SP.PDF 2018-01-11 10995-14657-1-SP.PDF 2018-01-11 10995-14658-1-SP.PDF 2018-01-11 10995-14659-1-SP.PDF 2018-01-11 10995-14660-1-SP.PDF 2018-01-11 10995-14661-1-SP.PDF 2018-01-11 10995-14663-1-SP.PDF 2018-01-11 10995-14664-1-SP.PDF 2018-01-11
Submitter	Mr Arbi Haza Nasution
Date submitted	January 11, 2018 - 08:50 PM
Section	Computer Science and Information Technology
Editor	Mokhtar Beldjehem (Review) Vicente Garcia Diaz (Review) Yin Liu (Review) Eugene Yu-Dong Zhang (Review)
Abstract Views	769

Status

Status	Published Vol 9, No 1: February 2019
Initiated	2018-09-09
Last modified	2020-07-14

Submission Metadata

Authors

Name	Arbi Haza Nasution
Affiliation	Kyoto University and Universitas Islam Riau
Country	Indonesia
Bio Statement	—

Principal contact for editorial correspondence.

Name	Yohei Murakami
Affiliation	Ritsumeikan University
Country	Japan
Bio Statement	—
Name	Toru Ishida
Affiliation	Kyoto University
Country	Japan
Bio Statement	—

Title and Abstract

Title	Generating similarity cluster of Indonesian languages with semi-supervised clustering
Abstract	Lexicostatistic and language similarity clusters are useful for computational linguistic researches that depends on language similarity or cognate recognition. Nevertheless, there are no published lexicostatistic/language similarity cluster of Indonesian ethnic languages available. We formulate an approach of creating language similarity clusters by utilizing ASJP database to generate the language similarity matrix, then generate the hierarchical clusters with complete linkage and mean linkage clustering, and further extract two stable clusters with high language similarities. We introduced an extended k-means clustering semi-supervised learning to evaluate the stability level of the hierarchical stable clusters being grouped together despite of changing the number of cluster. The higher the number of the trial, the more likely we can distinctly find the two hierarchical stable clusters in the generated k-clusters. However, for all five experiments, the stability level of the two hierarchical stable clusters is the highest on 5 clusters. Therefore, we take the 5 clusters as the best clusters of Indonesian ethnic languages. Finally, we plot the generated 5 clusters to a geographical map.

Indexing

Academic discipline and sub-disciplines	Computer and Informatics; Computational Linguistics; Artificial Intelligence
Keywords	lexicostatistic, language similarity, hierarchical clustering, k-means clustering, semi-supervised clustering
Language	en

Supporting Agencies

Agencies	Japan Society for the Promotion of Science (JSPS); Indonesia Endownment Fund for Education (LPDP)

OpenAIRE Specific Metadata

ProjectID	—

References

References	—

International Journal of Electrical and Computer Engineering (IJECE)

#10995 Review

Submission

Authors	Arbi Haza Nasution, Yohei Murakami, Toru Ishida
Title	Generating similarity cluster of Indonesian languages with semi-supervised clustering
Section	Computer Science and Information Technology
Editor	Mokhtar Beldjehem (Review) Vicente Garcia Diaz (Review) Yin Liu (Review) Eugene Yu-Dong Zhang (Review)

Peer Review

Round 1

Review Version	10995-14654-1-RV.TEX 2018-01-11
Initiated	2018-05-25
Last modified	2018-05-25
Uploaded file	None

Editor Decision

Decision	Accept Submission 2018-08-22
Notify Editor	Editor/Author Email Record 2018-07-09
Editor Version	None
Author Version	10995-22103-1-ED.PDF 2018-07-06 DELETE 10995-22103-2-ED.TEX 2018-07-06 DELETE
Upload Author Version	Choose File No file chosen Upload

International Journal of Electrical and Computer Engineering (IJECE)
p-ISSN 2088-8708, e-ISSN 2722-2578