ABSTRACT

The number of online media information is scattered of course the accuracy of information submitted to the user should be a priority. But often found errors in writing or spelling on news labels that make the dissemination of information, especially on the news portal is not well conveyed. To overcome this it takes correction on every wrong label to be the right label. However, further problems reappear when the label is done on a large amount of data, thereby increasing the use of time for the correct labeling. One way of correcting labels is to compare each label on an news portal with a dictionary as the correct label reference. And to undertake and accelerate the process of correcting the huge amount of data that can be done grouping labels. Recommended correction of Indonesian news portal labeling system is tribunnews.com with three frequently accessed categories national news, sports, and celebrities used as sample data. The method used to correct the label is to use the Damerau Levenshtein Distance algorithm by comparing each label to the data dictionary. And at the end of the Hierarchical Clustering to classify each label as a label recommendation. The results of this study produce clusters that display each label recommendation on the corrected label. The maximum Damerau Levenshtein Distance on the Hierarchical Distance that is limited to 2 is the correct labeling reference value for displaying relevant information and accuracy to the information. The final conclusion of this research is the use of Damerau Levenshtein Distance algorithm can help to correct the label and use Hierarchical Clustering to display the recommendation of the correction of labels with large amounts.

Keywords : *label, Damerau Levenshtein Distance, clustering, Hierarchical Clustering*