

Pengembangan sistem pendeteksian plagiarisme dengan menggunakan latent semantic analysis, learning vector quantization, dan self-organizing map = Development of automatic plagiarism detector using latent semantic analysis, learning vector quantization, and self organizing map

Paskalis Nandana Yestha Nabhastala, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20473164&lokasi=lokal>

Abstrak

Penelitian yang dilakukan berupa pengembangan terhadap sistem pendeteksi plagiarisme otomatis sehingga dapat menerapkan jaringan saraf tiruan Self-Organizing Map SOM untuk melakukan klasifikasi terhadap hasil keluaran Latent Semantic Analysis. SOM dipilih untuk melakukan klasifikasi karena algoritma ini tidak perlu melakukan supervisi pada proses pembelajarannya sehingga dapat secara otomatis menentukan tingkat plagiarisme antar paragraf yang tidak mudah ditentukan secara langsung oleh manusia. Selain itu dilakukan perbandingan akurasi penentuan tingkat plagiarisme yang dimiliki oleh sistem apabila hanya menggunakan LSA saja, penggunaan LSA dengan SOM, dan penggunaan LSA dengan Learning Vector Quantization LVQ.

Penggunaan SOM dan LVQ dilakukan untuk melakukan klasifikasi tingkat plagiarisme dari hasil keluaran LSA. Penentuan tingkat plagiarisme sudah cukup dilakukan apabila hanya menggunakan LSA saja, dimana sudah dapat menghasilkan tingkat akurasi paling tinggi yaitu 86,24. Namun, penggunaan SOM dengan jumlah kelas sebanyak 2 dengan 3 parameter memberikan rata-rata tingkat akurasi yang sedikit lebih rendah, yaitu sebesar 82,00. Sedangkan penggunaan LVQ dengan jumlah kelas sebanyak 2 dengan 3 parameter juga memberikan rata-rata tingkat akurasi yang sedikit lebih tinggi dibandingkan, yaitu sebesar 82,10.

This research has concern on deployment of neural network algorithm Self Organizing Map SOM in automatic plagiarism detector so it could be used to classify the output from Latent Semantic Analysis. SOM is chosen because it is an unsupervised neural network algorithm. With unsupervised neural network, it could determine the plagiarism level between paragraf automatically, which hard for human to determine it. Other than deployment of SOM, this research also focusses on the comparison of accuracy of the system if the system only deploys pure LSA, combination of LSA and SOM, and combination of LSA and Learning Vector Quantization LVQ.

SOM and LVQ are used to do classification for the output from LSA. Plagiarism level could be determined by the result of LSA only. It has 86,24 as the highest accuracy level. But, the usage of SOM with 2 classes and 3 parameters gives lower average of accuracy, which is 82,00 . Therefore, usage of LVQ with 2 classes and 3 parameters gives slight better accuracy than SOM, which is 82,10.