

Pengembangan algoritma pemrosesan paralel dengan openmp pada sistem pendeteksi plagiarisme dwibahasa berbasis latent semantic analysis dan learning vector quantization = Development of parallel processing algorithm with openmp on latent semantic analysis and learning vector quantization based bilingual plagiarism detection system

Ihsan Ibrahim, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20476091&lokasi=lokal>

Abstrak

Di Indonesia yang mayoritas karya tulis di dunia akademiknya masih menggunakan bahasa Indonesia dan referensi yang digunakan mayoritas berbahasa Inggris, memudahkan terjadinya tindak plagiarisme daripada penggunaan bahasa yang sama. Departemen Teknik Elektro telah mengembangkan sistem pendeteksi plagiarisme dwibahasa berbasis Latent Semantic Analysis LSA . Lamanya eksekusi, membuat paralelisme menjadi solusi untuk mengurangi waktu eksekusi dari sistem. Pada penelitian ini dilakukan pengembangan dengan pemrosesan paralel terhadap sistem dengan menggunakan OpenMP. Proses yang diparalelkan adalah, yaitu Singular Value Decomposition SVD, operasi-operasi matriks, dan proses Learning Vector Quantization LVQ dengan melakukan pada pendekatan loop-loop-nya.

Pada pengujian yang dilakukan, akurasi dari proses paralel memiliki konsistensi yang baik karena hasil yang sama dengan proses serial dan didapatkan peningkatan kecepatan eksekusi sistem sebesar 4-7,9 . Dengan fenomena pemrosesan paralel dengan menggunakan 1 thread memiliki waktu eksekusi yang lebih lambat daripada proses serial. Sedangkan saat menggunakan 2 thread dan 4 thread, didapatkan hasil yang lebih cepat daripada proses serial meskipun penggunaan 4 thread hanya berbeda sedikit atau cenderung sama dengan 2 thread. Hal ini disebabkan adanya overhead OpenMP yang terjadi saat pemrosesan paralel berjalan sebesar 20 , dan overhead MySQL yang membuat proses menjadi sangat lama karena besarnya yang mencapai 70 saat proses serial dan 50 pada proses paralel.

Majority of academic environment in Indonesia is still using Indonesian language and its references are in English. This condition led to ease the plagiarism acts when compared to same language environment. Due to this problem, Department of Electrical Engineering has developed bilingual plagiarism detection system based on Latent Semantic Analysis LSA . Parallelism becomes a solution to duration of execution problem. Development of parallel processing on the system with using OpenMP was conducted in this research. The parallelized processes were Singular Value Decomposition SVD , matrices operations, and Learning Vector Quantization LVQ with approach on loops.

In the testing process, accuracy of the parallel process had the same accuracy with the serial process. It is mean that the parallel process has good consistency. Then, the result of execution time has 4 7.9 of improvement compared to the serial one. There was a phenomenon that 1 thread of parallel process had worse performance than the serial process. Furthermore, use of 2 threads and 4 threads in the parallel process had a better execution time, even 4 threads is only slightly better or tend to be the same with 2 threads. These happened due to overhead presences. OpenMP overhead appeared at 20 when parallel executed and MySQL had more with 70 of system computation process in serial and 50 when executed in parallel.