

PROPOSAL
IMPLEMENTASI DAN ANALISIS KINERJA WORD COUNT PARALEL PADA
DOKUMEN TEKS BESAR MENGGUNAKAN OPENMP

MATA KULIAH KOMPUTASI PARALEL DAN TERDISTRIBUSI



Dosen Pengampu : Firdhaus Hari S A H, ST., M.Eng.

Disusun oleh :

Adinda Putri Nur Rhoqimah	2023061022
Hana Fithri Sabiila	2023061023
Xyla Syarifatuzzahra A	2024062006

PROGRAM STUDI INFORMATIKA
FAKULTAS SAINS TEKNOLOGI DAN KESEHATAN
UNIVERSITAS SAHID SURAKARTA
TAHUN AJARAN 2025/2026

1. Judul

Implementasi Word Count pada Dokumen Teks Besar Menggunakan Komputasi Paralel (Opsi B : Pengolahan Data Besar)

2. Deskripsi Masalah

Perkembangan teknologi informasi menyebabkan peningkatan jumlah data teks dalam skala besar, seperti dokumen digital, artikel daring, log sistem, dan dataset ensiklopedia digital. Salah satu proses dasar dalam pengolahan data teks adalah word count, yaitu menghitung jumlah kata dalam sebuah dokumen.

Pada dokumen berukuran kecil, proses word count dapat dilakukan dengan cepat menggunakan algoritma sekuensial. Namun, ketika ukuran dokumen mencapai ratusan megabyte, waktu eksekusi menjadi jauh lebih lama karena seluruh proses dilakukan oleh satu inti prosesor. Hal ini menimbulkan permasalahan efisiensi dan kinerja komputasi.

Oleh karena itu, diperlukan pendekatan komputasi paralel untuk mempercepat proses pengolahan data teks besar dengan memanfaatkan arsitektur multi-core yang tersedia pada komputer modern.

3. Justifikasi Pemilihan Masalah

Permasalahan word count dipilih karena merupakan contoh nyata dari pengolahan data besar yang sederhana secara konsep, namun menantang dari sisi kinerja ketika diterapkan pada dataset berukuran besar. Masalah ini sangat relevan untuk mempelajari dan menerapkan konsep paralelisasi data.

Selain itu, word count memiliki karakteristik yang cocok untuk komputasi paralel karena proses perhitungannya dapat dibagi ke beberapa bagian data dan dijalankan secara bersamaan. Dengan menggunakan OpenMP pada sistem shared memory, peningkatan performa dapat dianalisis secara kuantitatif melalui pengukuran waktu eksekusi, speedup, dan efficiency.

Pemilihan masalah ini juga sesuai sebagai studi kasus pembelajaran dalam mata kuliah Komputasi Paralel dan Terdistribusi.

4. Dataset / Input yang akan digunakan

Dataset yang digunakan berupa file teks berukuran besar yang bersumber dari kumpulan data Wikipedia.

Spesifikasi input:

- 1) Nama file: WikipediaDatabase.txt\
- 2) Jenis file: Text Document (.txt)
- 3) Ukuran file: ± 170 MB
- 4) Karakteristik data: Teks panjang berisi artikel ensiklopedia dengan jumlah kata yang sangat besar

Dataset ini dipilih karena memiliki ukuran dan kompleksitas yang cukup tinggi untuk menguji efektivitas algoritma word count pada skenario pengolahan data besar.

5. Output yang diharapkan

Output yang diharapkan dari proyek ini adalah:

- 1) Jumlah total kata dalam dokumen teks besar yang dihitung secara akurat.
- 2) Perbandingan kinerja antara implementasi sekuensial dan paralel.
- 3) Nilai execution time, speedup, dan efficiency untuk setiap variasi jumlah thread.
- 4) Kesimpulan mengenai efektivitas penerapan komputasi paralel pada permasalahan word count.