

Abstrak

BLAST (*Basic Local Alignment Search Tool*) adalah algoritma yang paling umum dipakai untuk pencocokan sekuensi rantai DNA dari specimen organik, NCBI merupakan salah satu badan yang menangani penyimpanan dataset dan mempergunakan algoritma tersebut, akan tetapi pada umumnya pencacahan akan memakan energi dan waktu komputasi yang cukup besar jika diaplikasikan dalam dataset dengan jumlah besar. melakukan optimasi algoritma yang bersifat universal dapat mengubah standarisasi dalam bioinformatika, oleh karena itu optimasi dilakukan di ranah perangkat keras, pemrosesan dan manajemen database. Pada penelitian ini dikaji penggunaan NCBI BLAST yang dijalankan di platform *Apache Spark* dengan manajemen pemrosesan dataset *Apache Hadoop* sebagai langkah optimasi pemrosesan pencacahan pencarian sekuensi data protein nukleotida dari *specimen organic* dari *query*. Dari hasil kajian ini menunjukkan waktu pemrosesan yang tidak terpengaruh dari banyaknya karakter pencarian, kemudian data sekuens yang acak dan rusak masih tetap dapat dicocokkan dengan baik.

Kata Kunci: *BLAST, Apache spark, Hadoop*