

Implementasi metode cafe untuk indexing dan retrieval database genomik: uji kasus penerapan panjang gram yang berbeda-besa

Christin

Deskripsi Dokumen: <http://lib.ui.ac.id/opac/ui/detail.jsp?id=125090&lokasi=lokal>

Abstrak

Database genomik semakin banyak digunakan oleh ahli biologi sebagai alat bantu utama mendapatkan informasi genetika dari suatu organisme melalui rangkaian DNA organisme tersebut. Hal yang biasanya dilakukan adalah mencari rangkaian DNA yang mempunyai susunan basa yang mirip dengan rangkaian DNA yang ingin dipelajari. Berbagai metode telah dikembangkan untuk mencari DNA-DNA dalam database genomik yang jumlahnya semakin meningkat. Metode-metode dikembangkan dengan pendekatan yang berbeda-beda, misalnya meningkatkan keakuratan dengan cara mengurangi kecepatan proses. Salah satu metode adalah metode CAFE yang menggunakan inverted indek yang merepresentasikan DNA-DNA dalam database genomik untuk mempercepat pencarian DNA. Metode ini pertama kali diperkenalkan oleh Hugh Williams pada tahun 1998. Tugas akhir ini mencoba mengimplementasikan metode CAFE untuk mencari DNA dan membandingkan DNA dalam database dengan DNA query. Tugas akhir ini akan mengevaluasi retrieval effectiveness dari metode CAFE yang menggunakan panjang gram yang berbeda-beda. Panjang gram yang kecil menyebabkan pencarian semakin teliti sehingga nilai recall semakin besar. Semakin panjang gram yang digunakan maka nilai precision akan semakin tinggi. Bertambahnya nilai threshold akan memperbesar nilai recall dan menurunkan nilai precision. Hasil eksperimen menunjukkan bahwa metode CAFE merupakan metode yang efektif karena dapat sebagian besar rangkaian yang relevan diakses pada setiap eksekusi query.