

Riassunto

La progettazione di metodi integrati di compressione e cifratura che garantiscano la riservatezza dei dati riducendone al contempo la dimensione è un argomento di ricerca molto attuale in Informatica.

D'altro canto, l'avvento delle nuove piattaforme di sequenziamento NGS e, più in generale, delle tecnologie cosiddette high-throughput ha consentito negli ultimi anni di ottenere velocemente, e a costi molto ridotti rispetto al passato, notevoli quantità di dati genomici.

Di conseguenza è sempre più sentita l'esigenza di costruire grandi database che integrino le informazioni genomiche con dati personali e clinici: database di questo genere permetterebbero, infatti, di ottenere risultati migliori nella cura di malattie multifattoriali e una comprensione più approfondita dei fenomeni biologici che avvengono nel nostro organismo; essi potrebbero costituire, tuttavia, un notevole problema per la privacy dei pazienti.

Mentre la confidenzialità dei dati clinici può essere ottenuta con tecniche di cifratura standard, comunemente applicate nel campo dei database relazionali, lo stesso non può dirsi per i dati genomici; il principale obiettivo del mio lavoro di ricerca è stato, quindi, la progettazione di nuovi schemi integrati di compressione e cifratura che consentissero di gestire dati genomici preservandone la riservatezza.

Affinchè tali sistemi fossero sfruttabili in applicazioni reali, ho profuso grande impegno nella progettazione di algoritmi e strutture-dati che consentissero di effettuare ricerche efficienti direttamente sui dati in forma compressa e cifrata, caricando di volta in volta in memoria RAM solo il minimo set di informazioni necessario ad eseguire ogni specifica operazione di ricerca.