

## Proposte di tesi e stage in ambito bio-informatico

Tra parentesi i docenti di riferimento dell'area bio che hanno fornito le idee. Possono candidarsi tutti gli studenti dei corsi di laurea triennale e magistrale dell'area informatica e bio-informatica.

[prof. Delledonne, dott. Giorgetti]

- 1) analisi e gestione di grandi moli di dati di deep sequencing mediante algoritmi paralleli su computer multicore (160 GB RAM, due Xeon 6 core upgradabile a 4)
- 2) organizzazione di pipeline per l'analisi dei dati provenienti da sequenziamento massivo parallelo --> possibilita' borsa di studio da pochi mesi a 2 anni
- 3) creazione di pipeline di tool per elaborazione parallela
- 4) interrogazione automatica di piu' motori di ricerca a partire da un'unica richiesta utente

[prof.ssa Torriani, dott.ssa Felis, dott. Fracchetti]

- 1) grid computing per l'utilizzo di un gruppo di PC per l'esecuzione di algoritmi paralleli; a differenza del cluster, il grid computing prevede che il PC svolga le normali funzioni e solo quando non utilizzato venga usato per l'esecuzione dell'algoritmo parallelo.

Requisiti: client per Win, Linux e Mac

necessita' di definire un modo particolare di scrivere programmi paralleli che tenga conto del fatto che i nodi non sono dedicati a tempo pieno

- 2) interrogazione automatica di piu' motori di ricerca a partire da un'unica richiesta utente
- 3) interrogazione automatica e periodica di piu' motori di ricerca per verificare se ci sono informazioni nuove
- 4) dati 2 o piu' genomi, identificazione delle porzioni simili con parametrizzazione sulla lunghezza della finestra di confronto

Requisiti: grado di somiglianza sfumato (non solo on/off) magari usando falsi colori

- 5) codifica su file congiunta di piu' genomi con possibilita' di link tra porzioni di genomi diversi per facilitare la navigazione e eventuale annotazione delle informazioni ottenute dall'attivita' (4)

Idee: la codifica puo' prevedere una forma di compressione (se ho porzioni simili di genoma si puo' fare una codifica a dizionario) ?

- 6) creazione di un tool per semplificare l'analisi della relazione tra presenza di geni nel genoma e via metabolica per la sintesi di proteine
- 7) creazione di un tool grafico per l'analisi e navigazione di genomi codificati

con l'attivit  (5)

8) ricerca bibliografica su tool commerciali e opensource per la genomica comparativa

[dott. Davide Quaglia]

1) creazione di una dispensa sulla gestione delle attivita' in un sistema distribuito  
(sia esso un multi-processore, un laboratorio analisi, un laboratorio di ricerca bio,  
un ufficio amministrativo, un'azienda agro-alimentare, ecc...)

Requisiti: occorre spiegare in maniera semplice ma esaustiva i concetti di precedenza,  
concorrenza, deadlock, mutua esclusione, tempo di servizio, throughput totale

2) creazione di un simulatore per la gestione delle attivita' in un sistema distribuito  
(sia esso un multi-processore, un laboratorio analisi, un laboratorio di ricerca bio,  
un ufficio amministrativo, un'azienda agro-alimentare, ecc...)

Requisiti: occorre modellare in maniera semplice ma esaustiva i concetti di precedenza,  
concorrenza, deadlock, mutua esclusione, tempo di servizio, throughput totale

3) utilizzo delle nozioni del punto 1) o del simulatore del punto 2) per introdurre il concetto  
di Qualita' del Servizio nel sistema (ad esempio per dare priorit  alle analisi del pronto  
soccorso su quelle di routine)