

Approfondimenti a scelta

(1) On the Representation of de Bruijn Graphs
(recomb14)

(2) Detecting Superbubbles in Assembly Graphs
(wabi13)

(3) PASTA: Ultra-Large Multiple Sequence Alignment
(recomb14)

(4) On the Complexity of Minimum Path Cover with
Subpath Constraints for Multi-assembly (recomb-seq)

(5) Relative FM-indexes (draft)

(6) Building a Pangenome Reference for a Population
(recomb14)

Preparare una presentazione di non più di 10 strutturata come segue.

Introduzione al problema biologico -motivazioni-: in che ambito biologico si colloca il problema? perché il lavoro è importante?

Il problema computazionale -definizione formale-: input-output, classificazione del problema (ottimizzazione, decisione), cosa c'è in letteratura (ci sono già studi/soluzioni di carattere computazionale relativi a questo problema)?, quali risultati originali vengono presentati da quel lavoro rispetto al problema trattato?

Complessità del problema e soluzione algoritmica: quali algoritmi vengono proposti? Si tratta di algoritmi euristici o esatti? Che tecnica algoritmica (ad es., Programmazione Dinamica, Greedy, etc.) viene impiegata? Come si classifica il problema rispetto alla complessità computazionale in tempo o spazio? Che strutture dati vengono utilizzate?

Sperimentazione: descrivere la sperimentazione, quali dati sono stati utilizzati e valutare i risultati prodotti. Come si comporta questo lavoro rispetto alla letteratura esistente?

Conclusioni: sintetizzare il contributo originale del lavoro in una riga. Come giudichi il lavoro letto in termini di originalità, chiarezza, organizzazione e presentazione a dei risultati?

Bibliografia: che referenze (citazioni) ritieni significative per capire l'articolo? La presentazione deve seguire lo schema della relazione. La relazione deve riportare esplicitamente i punti (a)-(f).

Implementazioni (gruppo da 3)

- Input: file GTF per gene, file FASTA della genomica
- Output:
 - lista introni del gene con il pattern dell'introne
 - tutti i trascritti del gene (formato FASTA)
 - tutte le CDS (formato FASTA)

Implementazioni (gruppo da 2)

- Input: file FASTQ di RNA-seq
- Output:
 - FASTA dei read filtrati sulla base di una certa qualità (trimming delle sequenze)