

# Esame del 30/06/2021 (Mattino)

Si consideri il database **genes\_small.sql**, estratto da un database del genoma umano originariamente creato per una challenge internazionale di biological data mining. I geni codificano le proteine che si localizzano in diverse parti della cellula e interagiscono fra di loro per eseguire le funzioni vitali della cellula stessa. Il dataset contiene informazioni sulla localizzazione (**tabella classification**), sulle interazioni (**tabella interactions**) e sulle funzioni dei geni (**tabella genes**), ed è strutturato secondo il diagramma ER illustrato nella pagina seguente. Il database è scaricabile dal repository: [https://github.com/TdP-datasets/genes\\_small](https://github.com/TdP-datasets/genes_small)

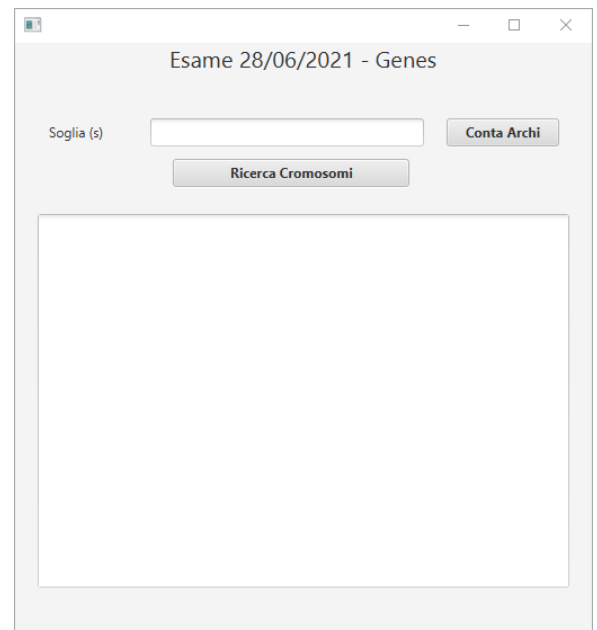
Si intende costruire un'applicazione JavaFX che permetta di svolgere le seguenti funzioni:

## PUNTO 1

- All'avvio del programma, si crei un grafo semplice, pesato e orientato i cui vertici siano tutti i cromosomi (tabella *genes*, colonna *chromosome*, considerando solo i valori *diversi da 0*).
- Un arco collega due cromosomi *diversi* solo se i due cromosomi contengono due geni (uno per cromosoma) che *compaiono* (nello stesso ordine) nella tabella *interactions*. Si noti che, per ciascun cromosoma, possono esistere più geni, e ciascuno di essi potrebbe essere presente più volte (associato a *function* diverse).
- Il peso di ciascun arco dovrà essere calcolato come la somma algebrica della correlazione (tabella *interactions*, campo *Expression\_Corr*), facendo attenzione a contare ogni coppia di geni *una sola volta*. Nell'esempio seguente, il peso dell'arco tra il cromosoma 5 ed il cromosoma 11 sarà pari a 2,468455212.

| Chromosome1 | Chromosome2 | Gene1   | Gene2   | Expression_Corr |
|-------------|-------------|---------|---------|-----------------|
| 5           | 11          | G234334 | G234684 | 0,567936255     |
| 5           | 11          | G235312 | G234684 | 0,709874329     |
| 5           | 11          | G237509 | G235586 | 0,937991552     |
| 5           | 11          | G238510 | G239467 | 0,252653076     |

- Si visualizzi il numero di vertici ed archi del grafo, ed i valori minimo e massimo dei pesi degli archi.
- Permettere all'utente di inserire un valore *soglia* (*S*), verificando che tale valore sia compreso nell'intervallo minimo-massimo calcolato al punto d).
- Alla pressione del bottone "Conta archi" stampare il numero di archi il cui peso è  $\leq S$ , ed il numero di archi il cui peso è  $> S$ .



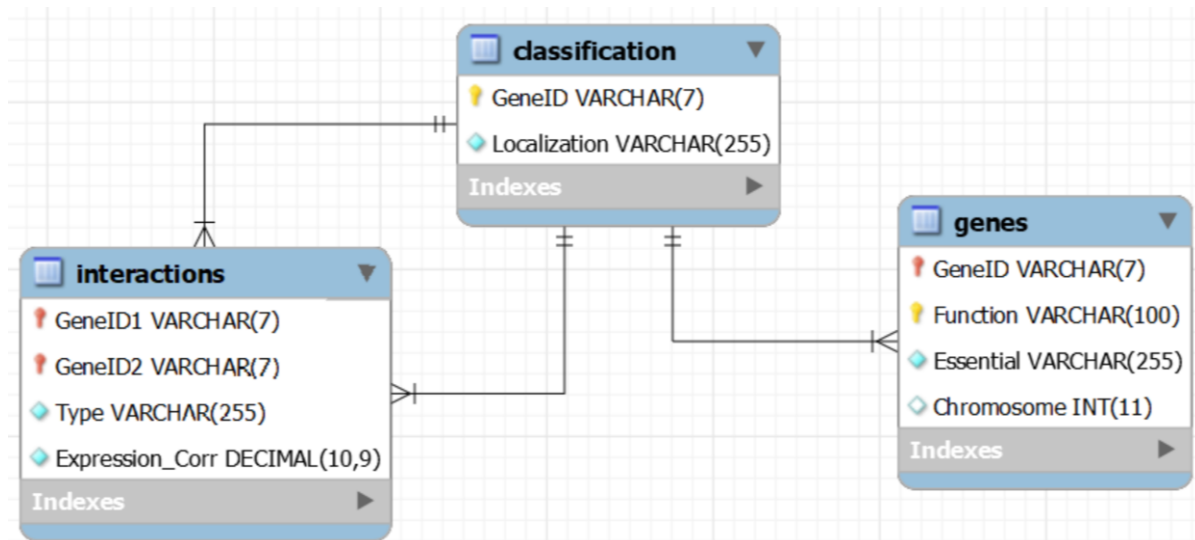
## PUNTO 2

- A partire dal grafo calcolato al punto precedente, alla pressione del tasto "Ricerca Cammino", si avvii una procedura di ricerca ricorsiva per determinare il più lungo cammino di vertici (cromosomi) che sia composto esclusivamente da archi di peso  $> S$ . La lunghezza del cammino sarà valutata dalla somma dei pesi degli archi incontrati.
- Si stampi la sequenza di cromosomi di lunghezza massima così ottenuta.

Nella realizzazione del codice, si lavori a partire dalle classi (Bean e DAO, FXML) e dal database contenuti nel progetto di base. È ovviamente permesso aggiungere o modificare classi e metodi.

Tutti i possibili errori di immissione, validazione dati, accesso al database, ed algoritmici devono essere gestiti, non sono ammesse eccezioni generate dal programma. Nelle pagine seguenti, sono disponibili due esempi di risultati per controllare la propria soluzione.

Le tabelle **classification**, **interactions** e **genes** contengono rispettivamente informazioni sulla localizzazione dei geni (e delle rispettive proteine), sul tipo e intensità del rapporto di interazione, e sulla funzione da loro espressa. Tutte le tabelle sono collegate tra loro tramite la chiave primaria che identifica i geni (GeneID).



## ESEMPI DI RISULTATI PER CONTROLLARE LA PROPRIA SOLUZIONE:

Esame 30/06/2021 (mattino) - Genes

Soglia (s)

Grafo creato: 16 vertici, 197 archi  
 Peso minimo = -1,268924, peso massimo = 7,661539

Esame 30/06/2021 (mattino) - Genes

Soglia (s)

Grafo creato: 16 vertici, 197 archi  
 Peso minimo = -1,268924, peso massimo = 7,661539  
 Soglia: 1,000000 --> Maggiori 91, minori 106  
 Soglia: 0,000000 --> Maggiori 164, minori 31  
 Soglia: -1,000000 --> Maggiori 196, minori 1  
 Soglia: 5,000000 --> Maggiori 4, minori 193