

# Esame del 30/06/2021 (Pomeriggio)

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)

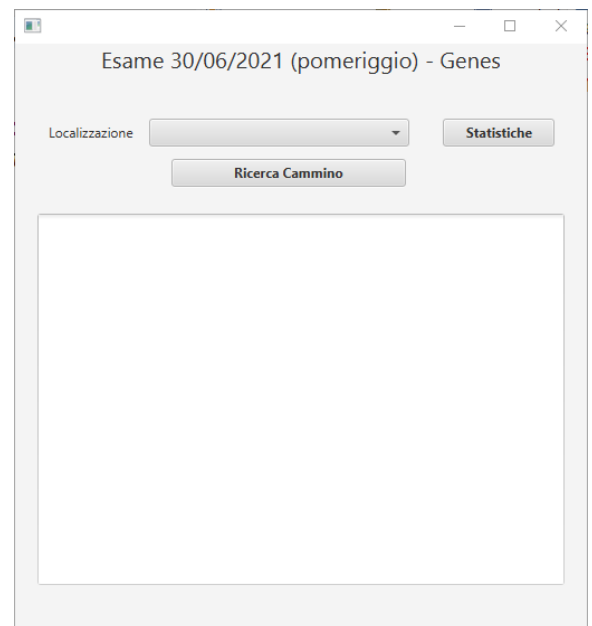
Si intende costruire un'applicazione JavaFX che permetta di svolgere le seguenti funzioni:

## PUNTO 1

- All'avvio del programma, si crei un grafo semplice, pesato e non orientato i cui vertici siano tutte le localizzazioni (tabella *classification*, colonna *localization*).
- Ogni localizzazione contiene un insieme di geni, i quali possono essere collegati tra di loro attraverso la tabella *interactions*. **I vertici corrispondenti a due localizzazioni saranno collegati da un arco se e solo se esiste almeno una interazione che coinvolge due geni**, rispettivamente della prima della seconda localizzazione (o nell'ordine inverso).
- Il peso di ciascun arco dovrà essere un numero intero, pari al numero di tipi<sup>1</sup> diversi di interazioni tra i geni associati alle due localizzazioni. Nell'esempio seguente (che riporta solo un sottoinsieme delle righe), vi sono molti geni associati a cytoplasm che interagiscono con geni associati a cytoskeleton, ma i tipi di interazione riscontrati in questo caso sono 3.

Localization1	Localization2	Gene1	Gene2	Type
cytoplasm	cytoskeleton	G235632	G234392	Genetic
cytoplasm	cytoskeleton	G239673	G234392	Genetic
cytoplasm	cytoskeleton	G235301	G234732	Genetic-Physical
cytoplasm	cytoskeleton	G235407	G234732	Genetic-Physical
cytoplasm	cytoskeleton	G235477	G234732	Physical
cytoplasm	cytoskeleton	G235632	G234732	Genetic-Physical
cytoplasm	cytoskeleton	G235687	G234732	Genetic-Physical
cytoplasm	cytoskeleton	G235632	G234732	Genetic

- Si visualizzi il numero di vertici ed archi del grafo.
- Permettere all'utente di selezionare una localizzazione, ed alla pressione del tasto "Statistiche", stampare tutte le localizzazioni connesse a quelle selezionata, visualizzando per ciascuna di esse il peso dell'arco corrispondente.



## PUNTO 2

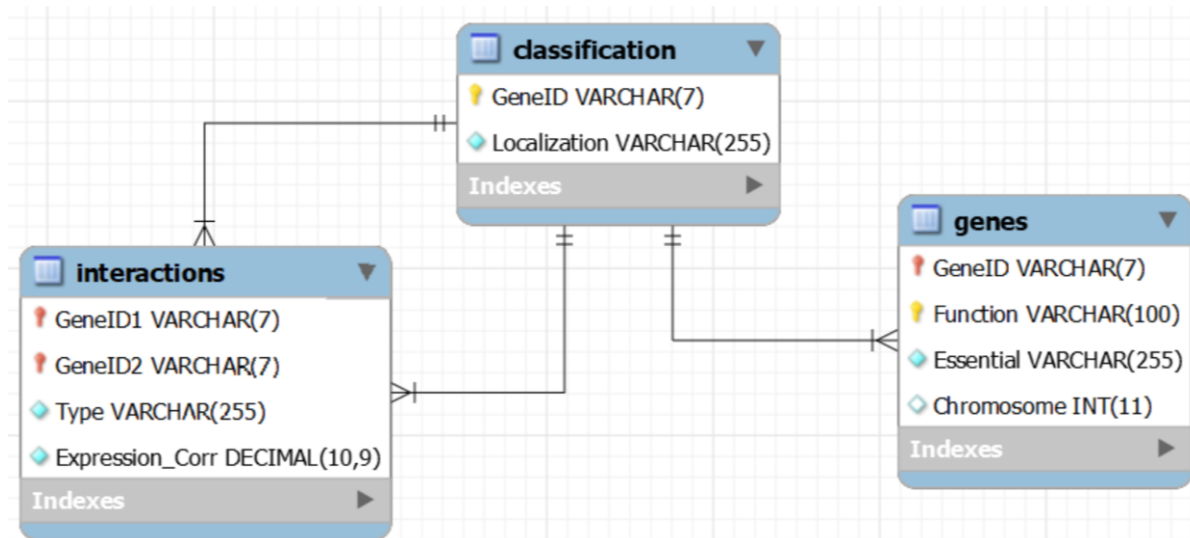
- A partire dal grafo calcolato al punto precedente, alla pressione del tasto "Ricerca Cammino", si avvia una procedura di ricerca ricorsiva per determinare il più lungo cammino semplice di localizzazioni che parta dalla localizzazione selezionata dall'utente. La lunghezza del cammino sarà valutata dalla somma dei pesi degli archi incontrati.
- Si stampi la sequenza di localizzazioni di lunghezza massima così ottenuta.

<sup>1</sup> **Attenzione** - Type è una parola riservata in SQL, per utilizzarlo come nome di colonna è necessario racchiuderlo tra apici inversi: ``Type``

Nella realizzazione del codice, si lavora 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:

