TdP – Esame del 18/07/2024 – Traccia A

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>

Si intende costruire un’applicazione FLET che permetta di svolgere le seguenti funzioni:

## PUNTO 1

1. Si permetta all’utente di selezionare due interi *a,b* che rappresentino i possibili valori numerici della variabile *Cromosoma* (tabella *genes*).
2. Si crei un grafo semplice, pesato e non orientato. I nodi sono tutti i geni contenuti nella tabella *genes* che appartengono ad un cromosoma nell’intervallo *a,b* fornito dall’utente. Un arco collega due geni *diversi* solo se i due geni presentano la stessa Localizzazione (tabella *classification*). Il peso dell’arco corrisponde all’indice di correlazione dell’interazione fra i due geni (tabella *interactions*). Il verso dell’arco sarà uscente dal gene con Cromosoma minore ed entrante nel gene con Cromosoma maggiore. Il caso in cui due geni appartengono allo stesso cromosoma va gestito aggiungendo entrambi gli archi.
3. Alla creazione del grafo, stampare automaticamente il numero di vertici ed archi del grafo.
4. Permettere all’utente di selezionare da un dropdown la lista delle localizzazioni disponibili nel database, e alla pressione del tasto *Dettagli* stampare tutti i geni che condividono la localizzazione scelta.

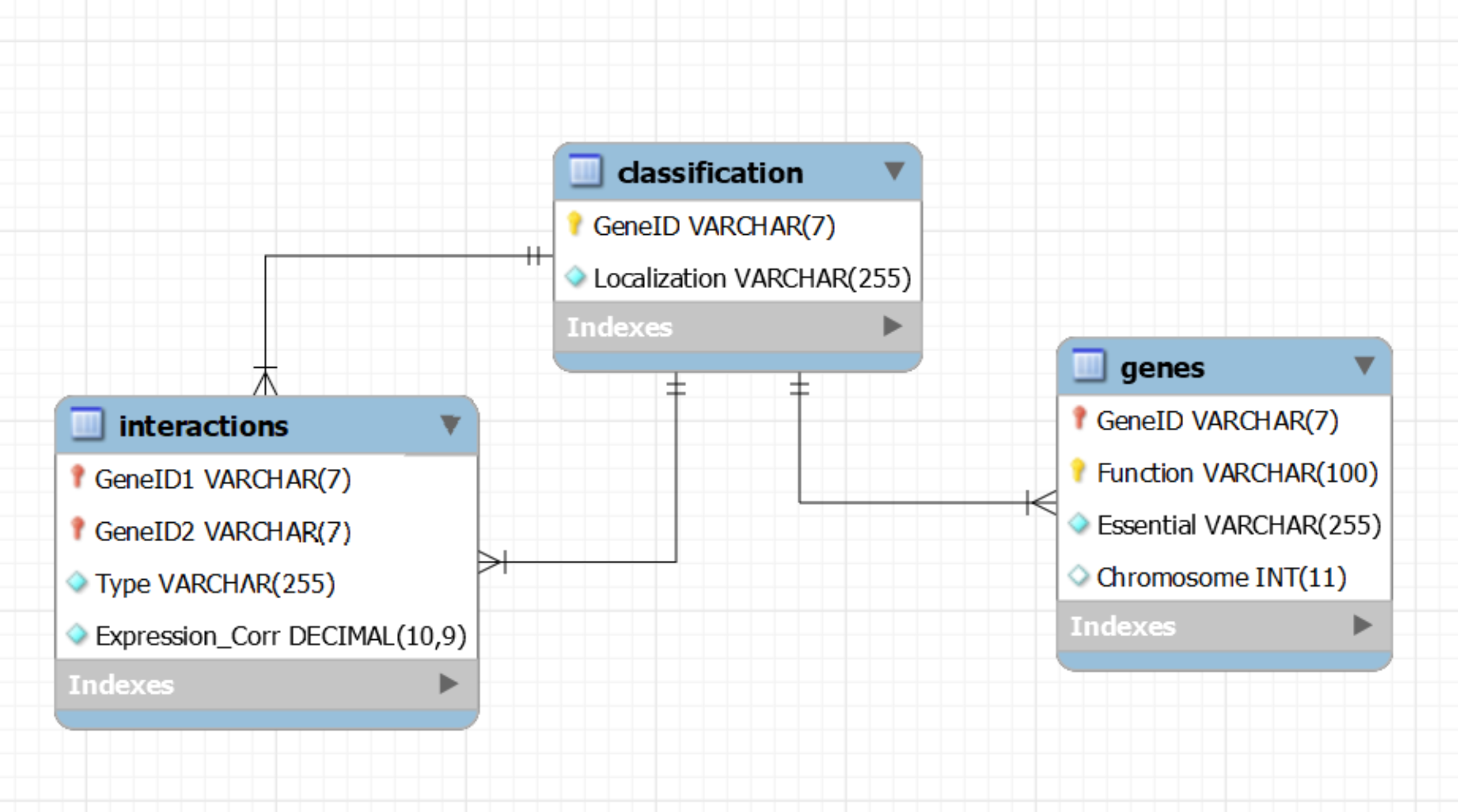
## PUNTO 2

1. A partire dal grafo definito al punto 1, si vuole implementare una procedura ricorsiva che identifichi il cammino più lungo che minimizza la somma dei pesi del percorso. Nello specifico, la procedura dovrà identificare il cammino più lungo ammissibile. In seconda istanza, tra i diversi cammini di pari lunghezza, dovrà prediligere il percorso a peso minimo.
2. Si stampi la sequenza di cromosomi di lunghezza massima così ottenuta.

Nella realizzazione del codice, si lavori a partire dalle classi 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:**