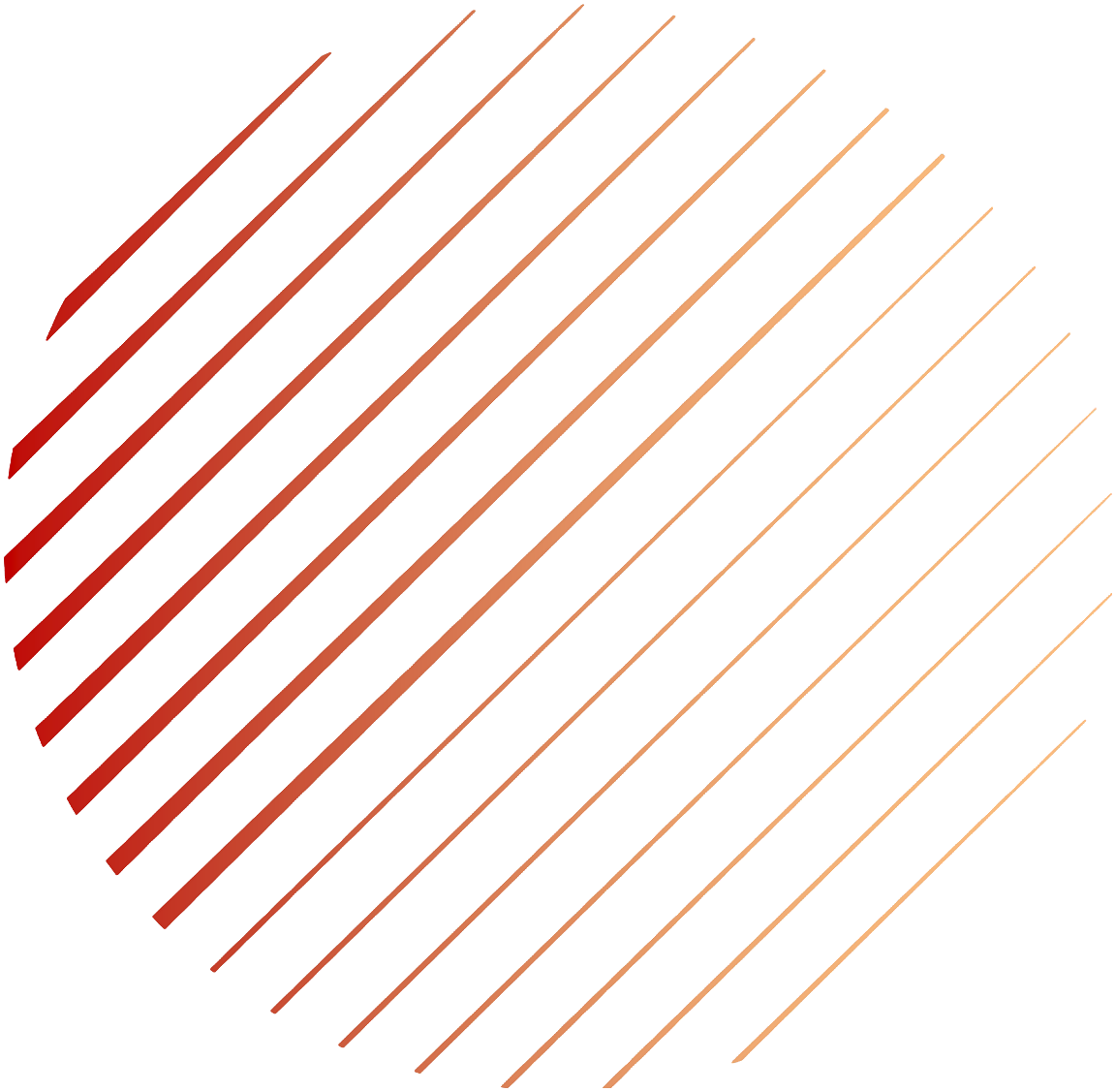


H-clus clustering

Caso di studio di Metodi Avanzati di Programmazione
AA 2022-2023



Realizzato Da

Ivan Digioia 716685

i.digioia3@studenti.uniba.it

SOMMARIO

1. INTRODUZIONE.....	3
2. INTRODUZIONE AL PROGETTO	5
3 DIAGRAMMI UML	6
4 GUIDA ALL' INSTALLAZIONE	14
5. GUIDA UTENTE.....	15

1. INTRODUZIONE

1.1 L'algoritmo H-Clus

H-Clus è un algoritmo di clustering gerarchico progettato per creare una struttura ad albero (dendrogramma) che rappresenta la gerarchia dei cluster nei dati. A differenza degli algoritmi di clustering basati su partizioni, come K-Means, H-Clus non richiede di specificare a priori il numero di cluster e permette di visualizzare la struttura gerarchica delle relazioni tra i dati.

Origine e Caratteristiche

H-Clus è stato sviluppato per gestire dati complessi, con lo scopo di individuare relazioni gerarchiche tra le osservazioni. Questo algoritmo costruisce una gerarchia di cluster, rappresentando i dati come un albero dove ogni nodo corrisponde a un cluster e i nodi foglia rappresentano i singoli dati. H-Clus può essere utilizzato sia in modalità agglomerativa (bottom-up) che divisiva (top-down), a seconda del contesto e degli obiettivi dell'analisi.

1.2 Funzionamento dell'algoritmo

L'algoritmo H-Clus si basa principalmente sull'approccio gerarchico agglomerativo, che funziona attraverso le seguenti fasi:

1. **Inizializzazione dei Cluster:** Ogni elemento del dataset inizia come un singolo cluster individuale. La distanza tra ciascun elemento viene calcolata utilizzando metriche come la distanza euclidea, di Manhattan o altre metriche di similarità (nel progetto sarà implementata solo la distanza euclidea).
2. **Fusione dei Cluster:** In ogni iterazione, H-Clus unisce i due cluster più vicini sulla base della distanza minima. Questo processo continua fino a quando tutti gli elementi sono raggruppati in un unico cluster globale. Durante la fusione, l'algoritmo aggiorna le distanze tra i cluster utilizzando metodi come il single o average link distance.
3. **Creazione del Dendrogramma:** Durante il processo di fusione, H-Clus costruisce un dendrogramma, un grafico che rappresenta la gerarchia di tutti i cluster. Gli utenti possono scegliere il livello di taglio dell'albero per identificare il numero di cluster ottimale in base alle esigenze dell'analisi.
4. **Determinazione dei Cluster Finali:** Una volta completato il dendrogramma, l'utente può selezionare il livello appropriato di profondità dell'albero per suddividere i dati in gruppi distinti.

1.3 Limiti

Complessità Computazionale: La costruzione del dendrogramma può essere computazionalmente intensiva per dataset molto grandi, rendendo H-Clus meno adatto per big data rispetto a metodi più scalabili.

Sensibilità alle Scelte di Distanza e Linkage: I risultati dell'H-Clus possono variare significativamente a seconda della scelta delle metriche di distanza e del metodo di linkage

2.INTRODUZIONE AL PROGETTO

2.1 Descrizione del progetto

Il software realizzato utilizza l'algoritmo H-Clus, descritto nella sezione precedente, esso elabora dati da una tabella presente in un database di tipo MySQL.

Il progetto, risultato di esercitazioni, consiste in un'applicazione di tipo Client/Server.

Il server si occupa di ricevere le richieste di un client, il quale può effettuare le seguenti operazioni:

- Generare un dendrogramma partendo dai dati del database e dagli inserimenti dell'utente come la profondità e la scelta tra single o average link distance e memorizza il risultato in un file.
- Caricare da un file il dendrogramma memorizzato, si preferiscono i file in estensione '.dat' ma andrà bene un qualsiasi file di letture come un '.txt'.

In entrambi i casi, il client dovrà specificare nei criteri di ricerca:

- la profondità per suddividere i dati.
- se operare il single o average link distance.
- il nome del file su cui salvare o caricare i dati.

L'interfaccia grafica è stata sviluppata usando la tecnologia JavaFX; inoltre, è stato utilizzato SceneBuilder per la creazione dell'interfaccia grafica e CSS.

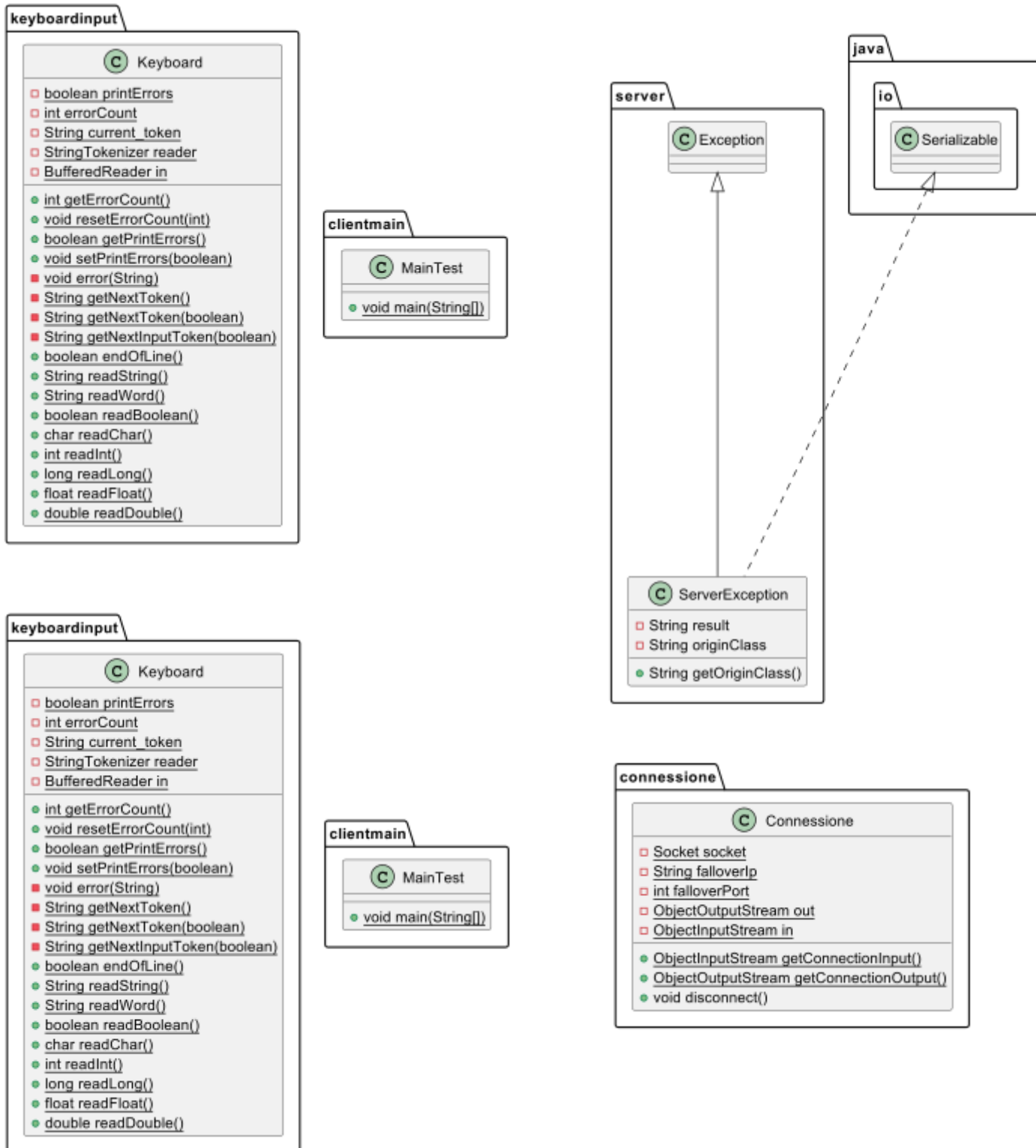
Nel progetto sono presenti entrambe le versioni, sia quella fruibile attraverso console, sia quella utilizzabile con l'interfaccia grafica. Nel presente documento verrà trattata solo l'estensione.

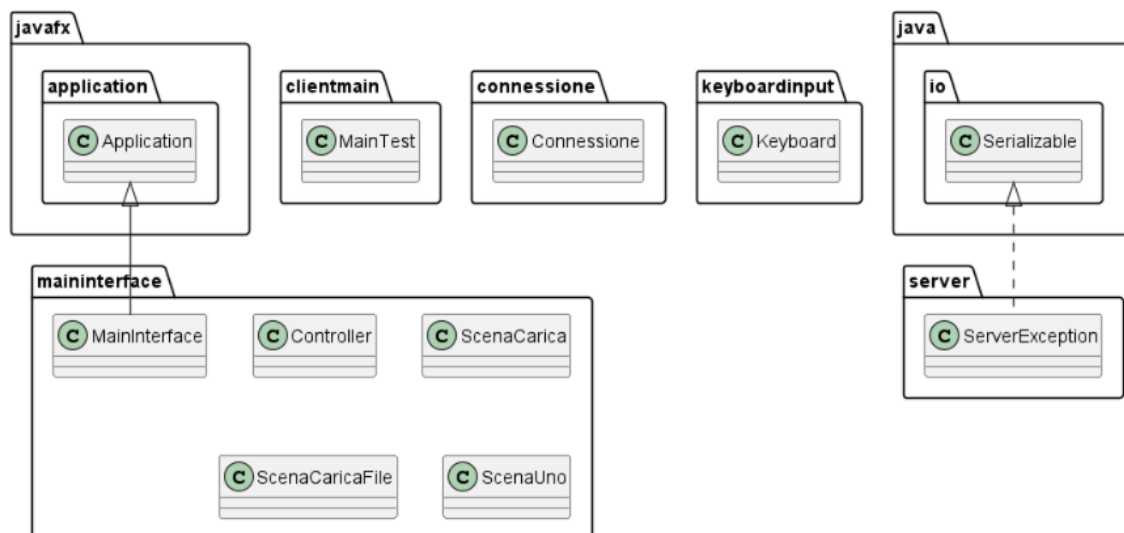
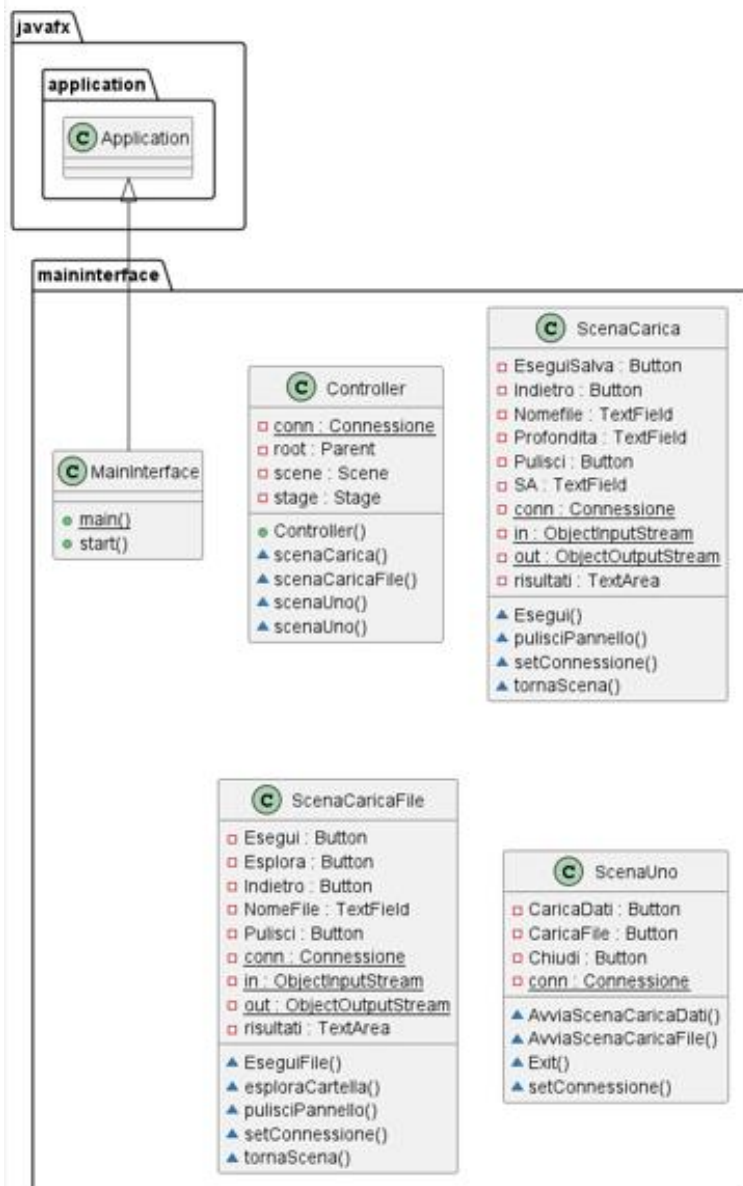
Nella sezione 3 sono riportati anche i diagrammi UML per il client e per il server. Inoltre, nella cartella "Javadoc" è stata allegata la Javadoc creata direttamente dall'IDE di sviluppo (IntelliJ). Nella sezione 5 del documento sono riportati esempi di esecuzione.

3 DIAGRAMMI UML

Segue la realizzazione dei diagrammi per la versione Base del MeansServer e KMeansClient

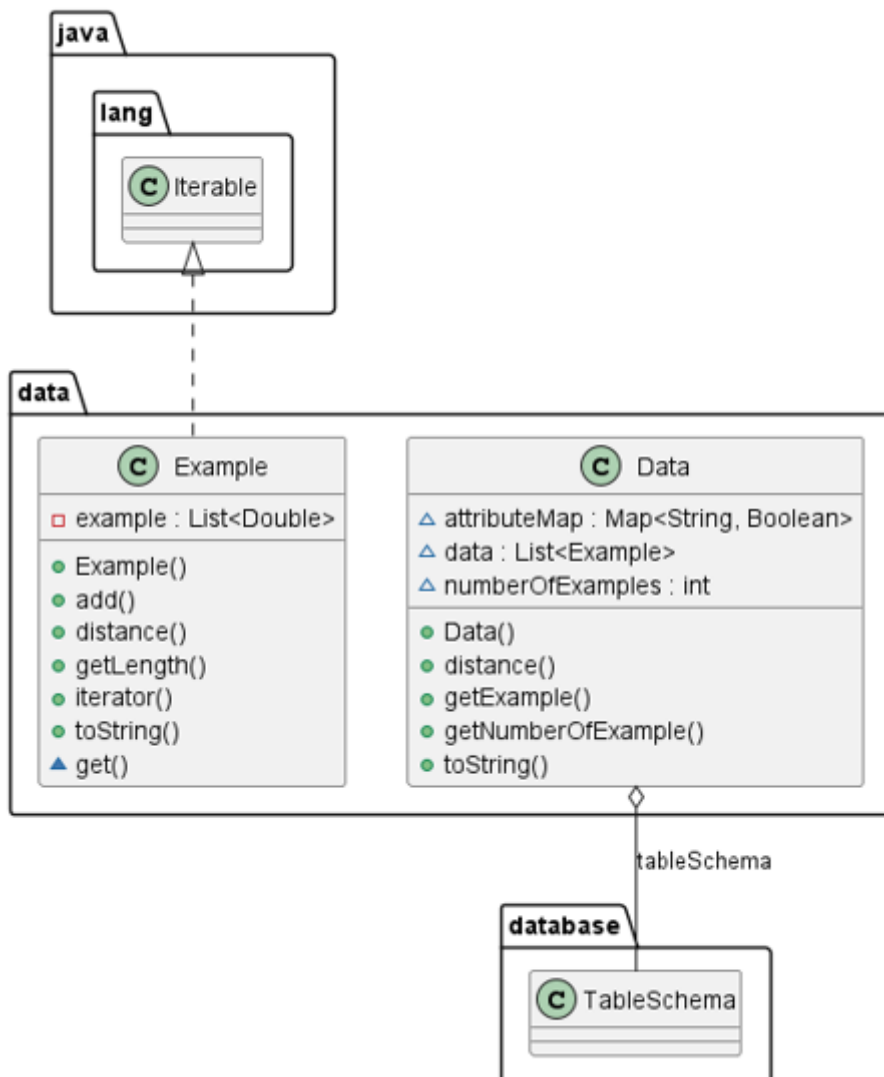
3.1 Client UML



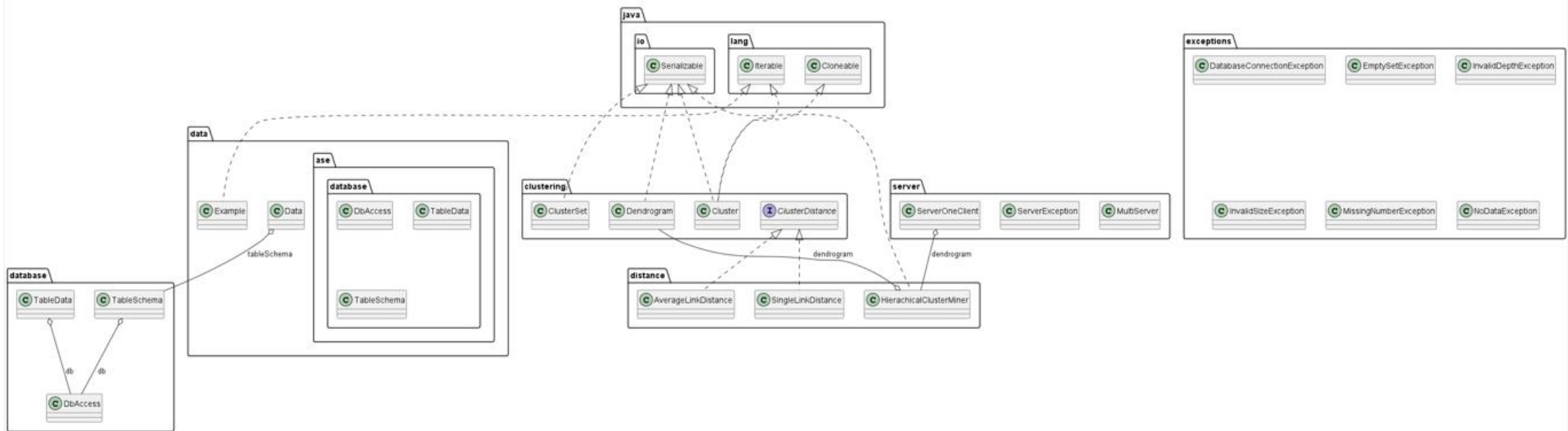
HCLUSCLIENT's Class Diagram**MAININTERFACE's Class Diagram**

2.2 Server UML

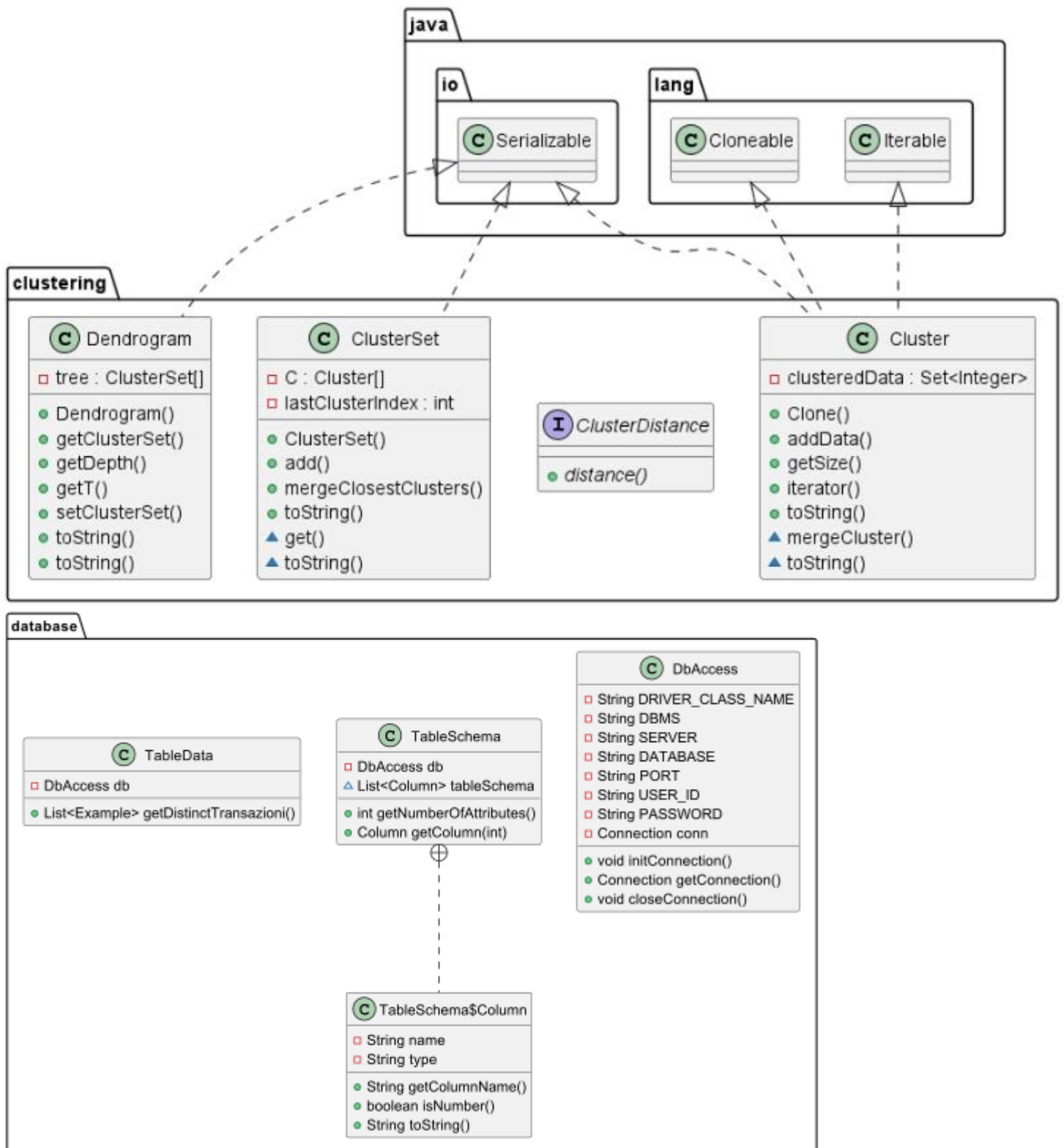
DATA's Class Diagram

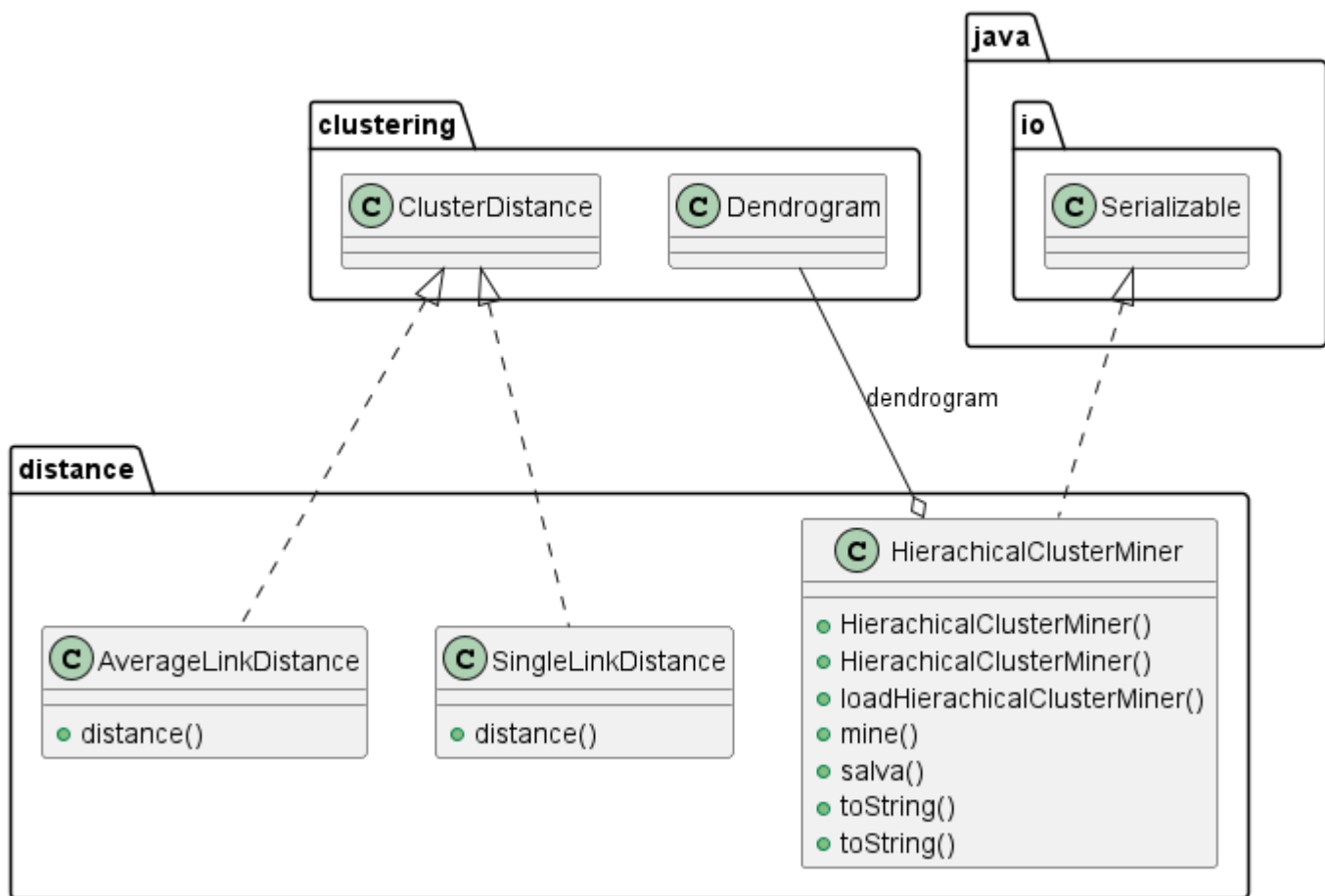
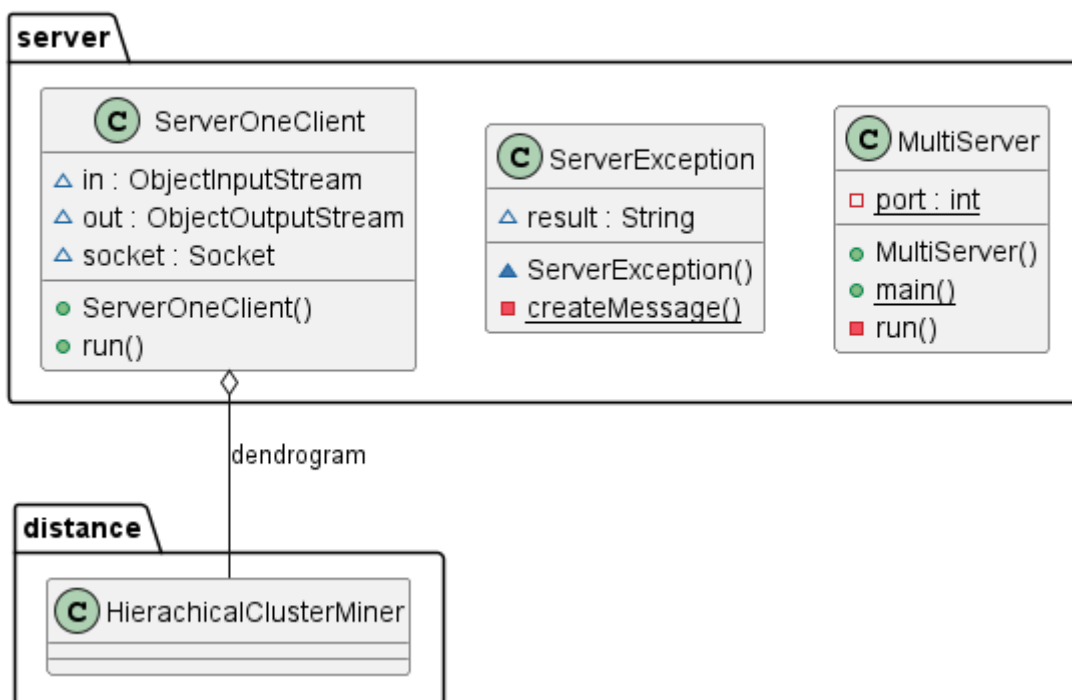


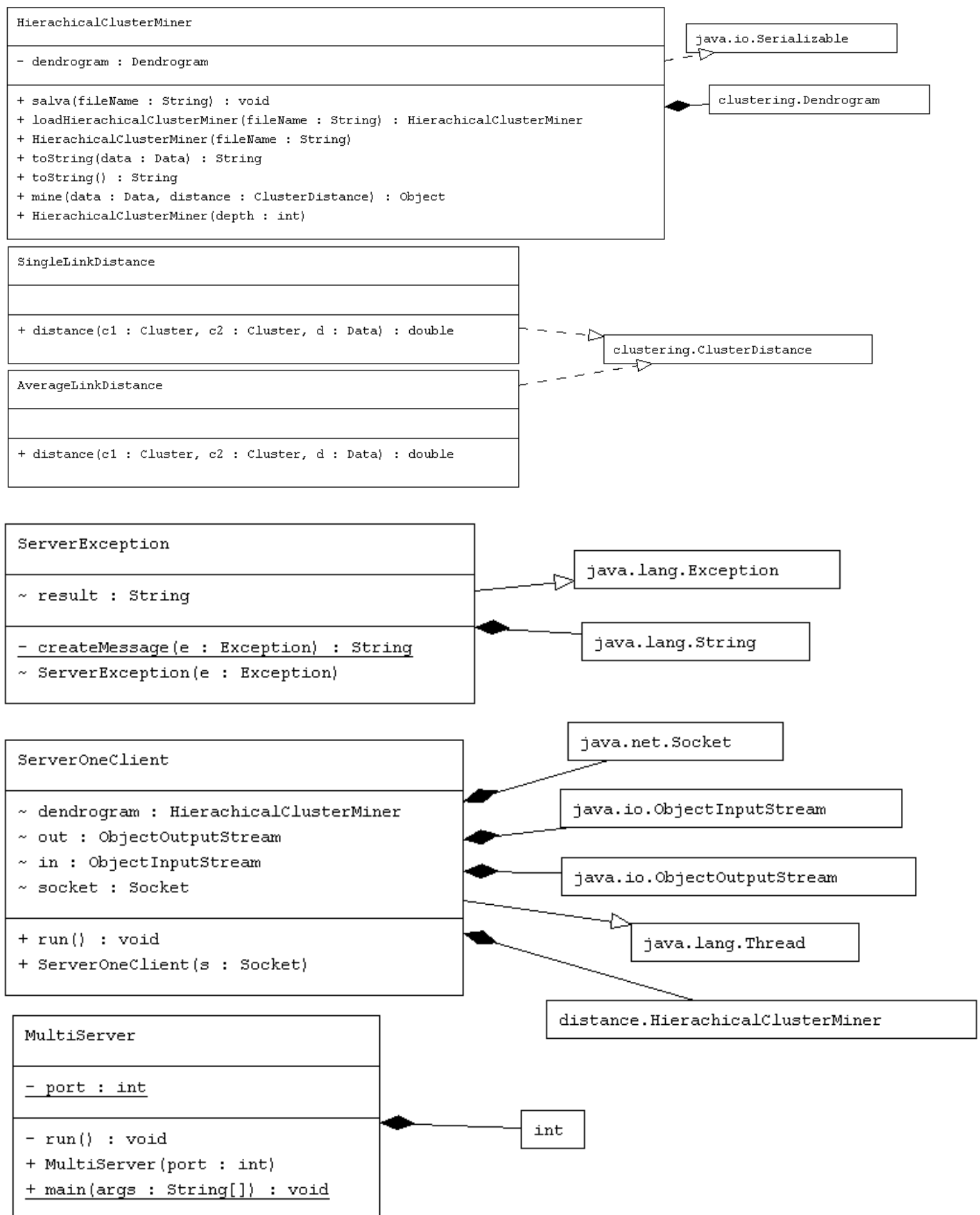
HCLUSERVER's Class Diagram

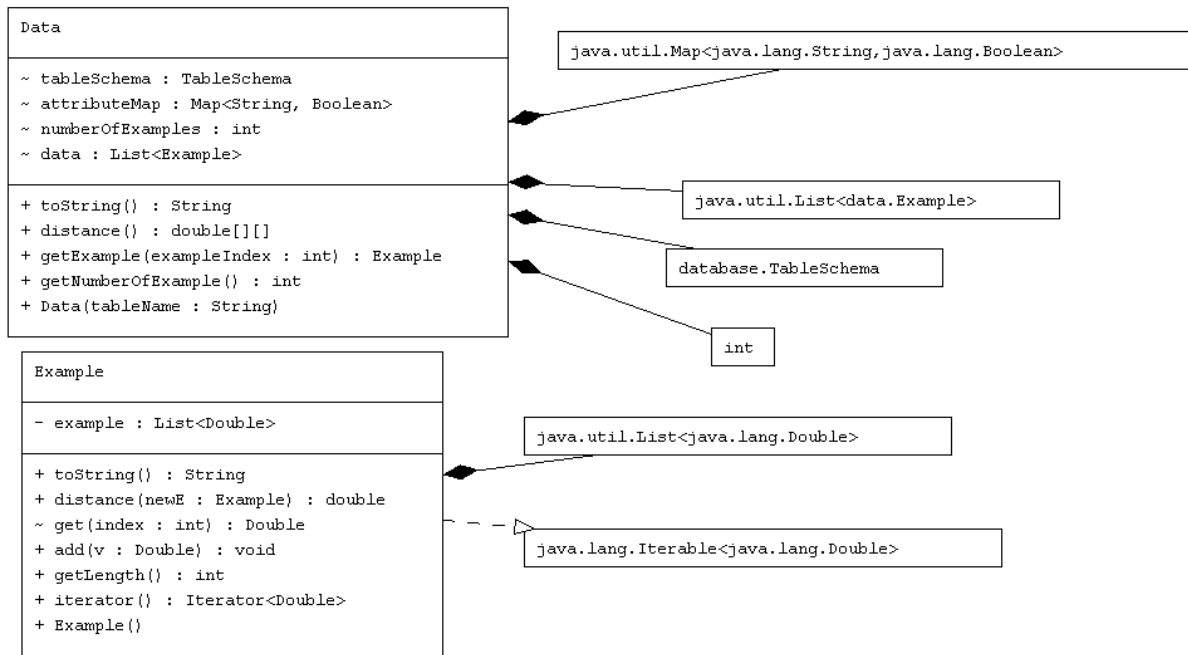


CLUSTERING's Class Diagram



DISTANCE's Class Diagram**SERVER's Class Diagram**





4 GUIDA ALL' INSTALLAZIONE

4.1 Installazione Server

Per il corretto funzionamento del progetto lato server è necessario:

- Spostare l'intera cartella del progetto sul desktop;
- Installare MySQL 8.0;
- Installare Java Runtime Environment (JRE) versione 20;
- Avviare il server MySQL;
- Eseguire lo script MySQL presente nella cartella "SQL Connector". Tale script inizializza il database con tabelle e tuple di esempio.ⁱ

Per avviare il client è possibile aprire il file *Eseguibile Client.bat* contenuto nella cartella "*Eseguibile/Estesa*". Alternativamente, è possibile avviare il client tramite riga di comando indicando (partendo dalla cartella in cui si trova il file

Eseguibile/Esteso/HclusClient.jar):

- La directory in cui è contenuto il java.exe (se non è contenuto nel PATH) - Il comando -jar che indica di avviare un file .jar La riga sarà simile a: C:\\$PathTo\$java.exe -jar HclusServer.jar

4.2 Installazione Client

Per il corretto funzionamento del progetto lato client è necessario:

- Installare Java Runtime Environment (JRE) versione 20;
- Avviare il server.ⁱⁱⁱⁱ

Per avviare il client è possibile aprire il file *Eseguibile Client.bat* contenuto nella cartella "*Eseguibile/Estesa*". Alternativamente, è possibile avviare il client tramite riga di comando indicando (partendo dalla cartella in cui si trova il file

Eseguibile/Esteso/HclusClient.jar):

- La directory in cui è contenuto il java.exe (se non è contenuto nel PATH);
- Il comando -jar che indica di avviare un file .jar;
- L'indirizzo IP a cui è collegato il server (di default 127.0.0.1);
- La porta su cui è in ascolto il server (di default 8080);
- Le librerie javaFx necessarie all'avvio del programma (base, controls, graphics, media, fxml);

La riga sarà simile a:

C:\\$PathTo\$java.exe -jar ".\lib" -add-modules

javafx.base,javafx.controls,javafx.graphics,javafx.media,javafx.fxml -jar HclusClient.jar 127.0.0.1 8080

5. GUIDA UTENTE

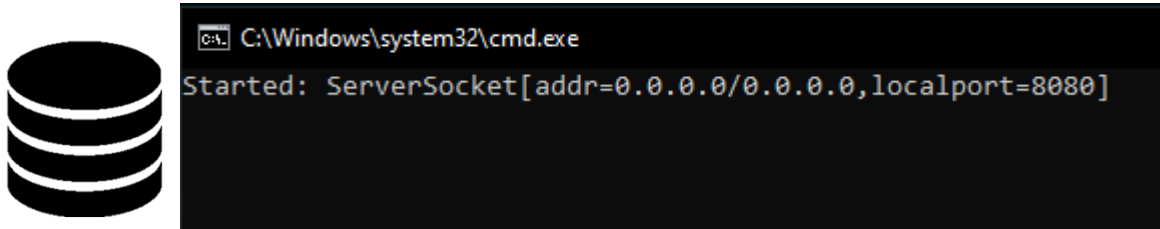
Nella cartella principale del progetto è presente una sottocartella *“File memorizzati”*, nella quale verranno salvati (e caricati) in file. In essa sono presenti già dei file a scopo di esempio.

La tabella di esempio presenti nello script MySQL si chiama *“exampletab”*.

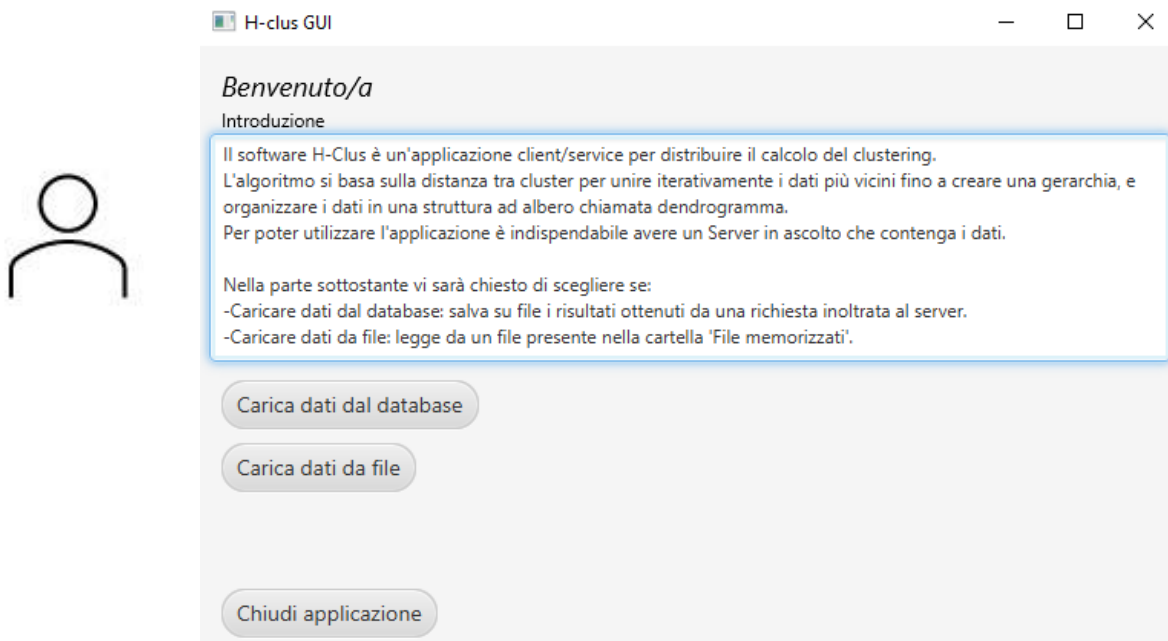
4.3 Guida all'interfaccia grafica

Nella cartella *“Eseguibile”* eseguire il file *“Eseguibile generale Esteso.bat”*. Si aprirà una schermata a linea di comando essa sarà il server che si avvia e successivamente si aprirà la prima schermata dell'interfaccia.

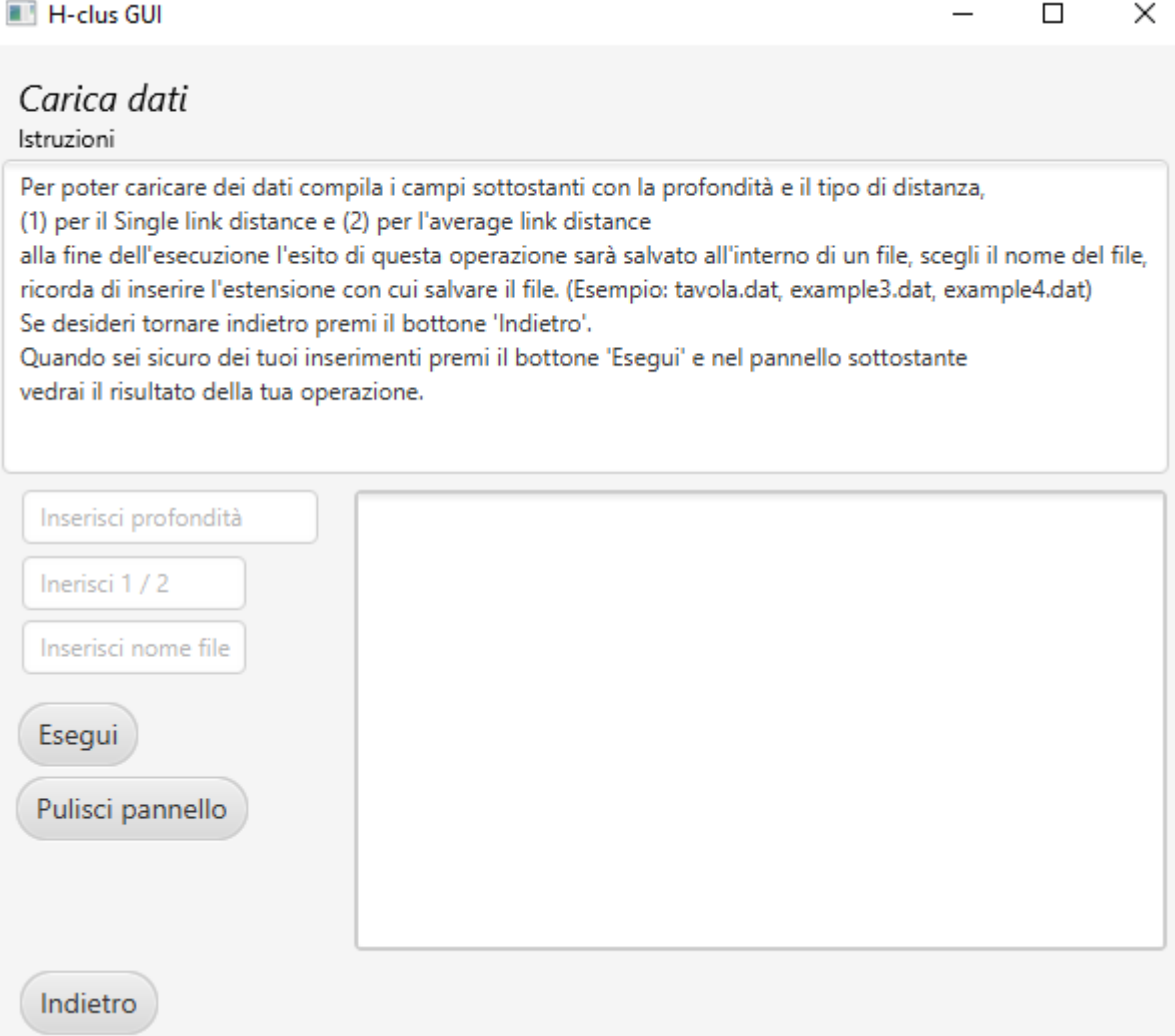
1) Avvio server:



2) Avvio Client:



3) Carica dati dal database:



Carica dati
Istruzioni

Per poter caricare dei dati compila i campi sottostanti con la profondità e il tipo di distanza, (1) per il Single link distance e (2) per l'average link distance alla fine dell'esecuzione l'esito di questa operazione sarà salvato all'interno di un file, scegli il nome del file, ricorda di inserire l'estensione con cui salvare il file. (Esempio: tavola.dat, example3.dat, example4.dat) Se desideri tornare indietro premi il bottone 'Indietro'. Quando sei sicuro dei tuoi inserimenti premi il bottone 'Esegui' e nel pannello sottostante vedrai il risultato della tua operazione.

Inserisci profondità

Inserisci 1 / 2

Inserisci nome file

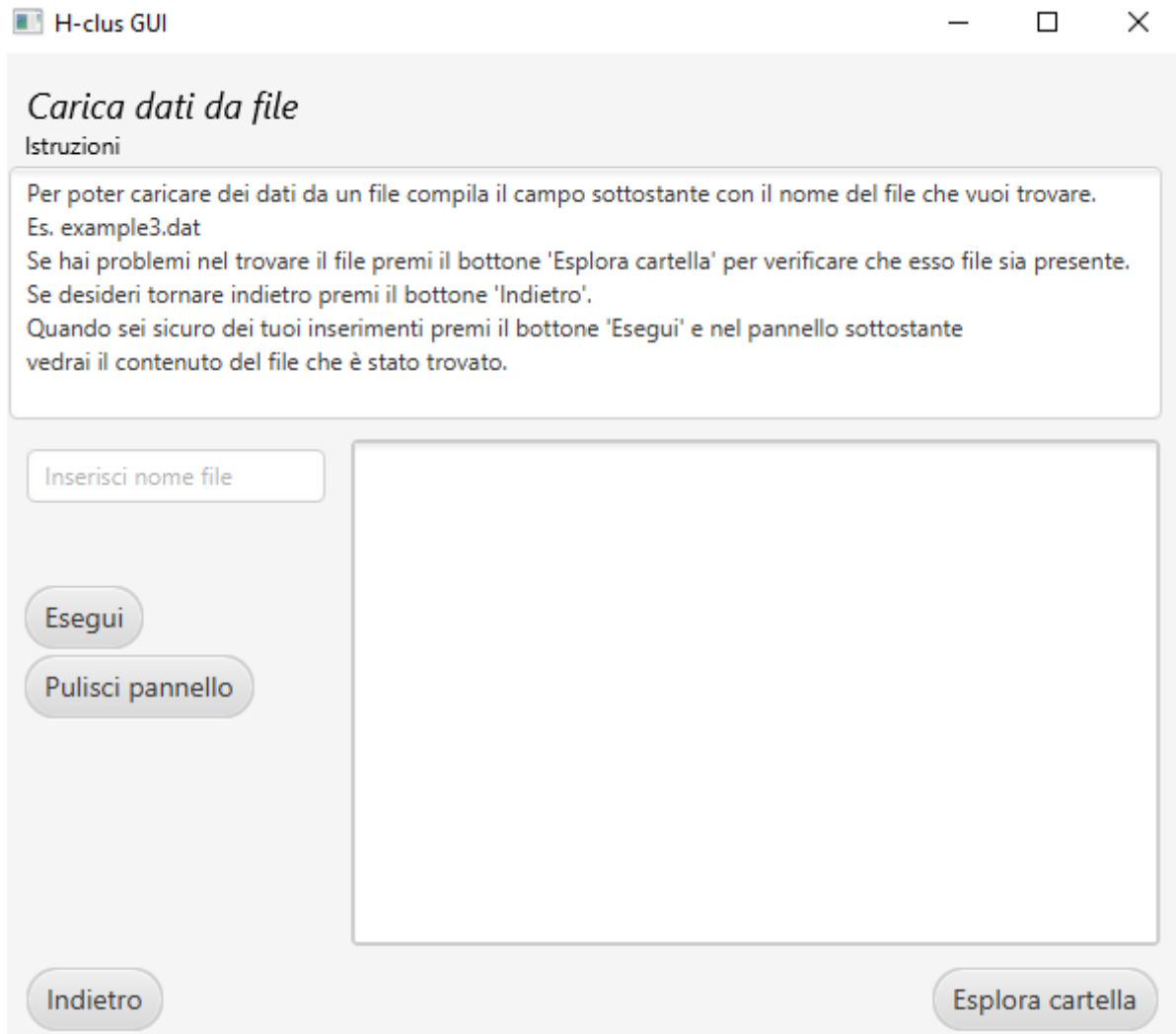
Esegui

Pulisci pannello

Indietro

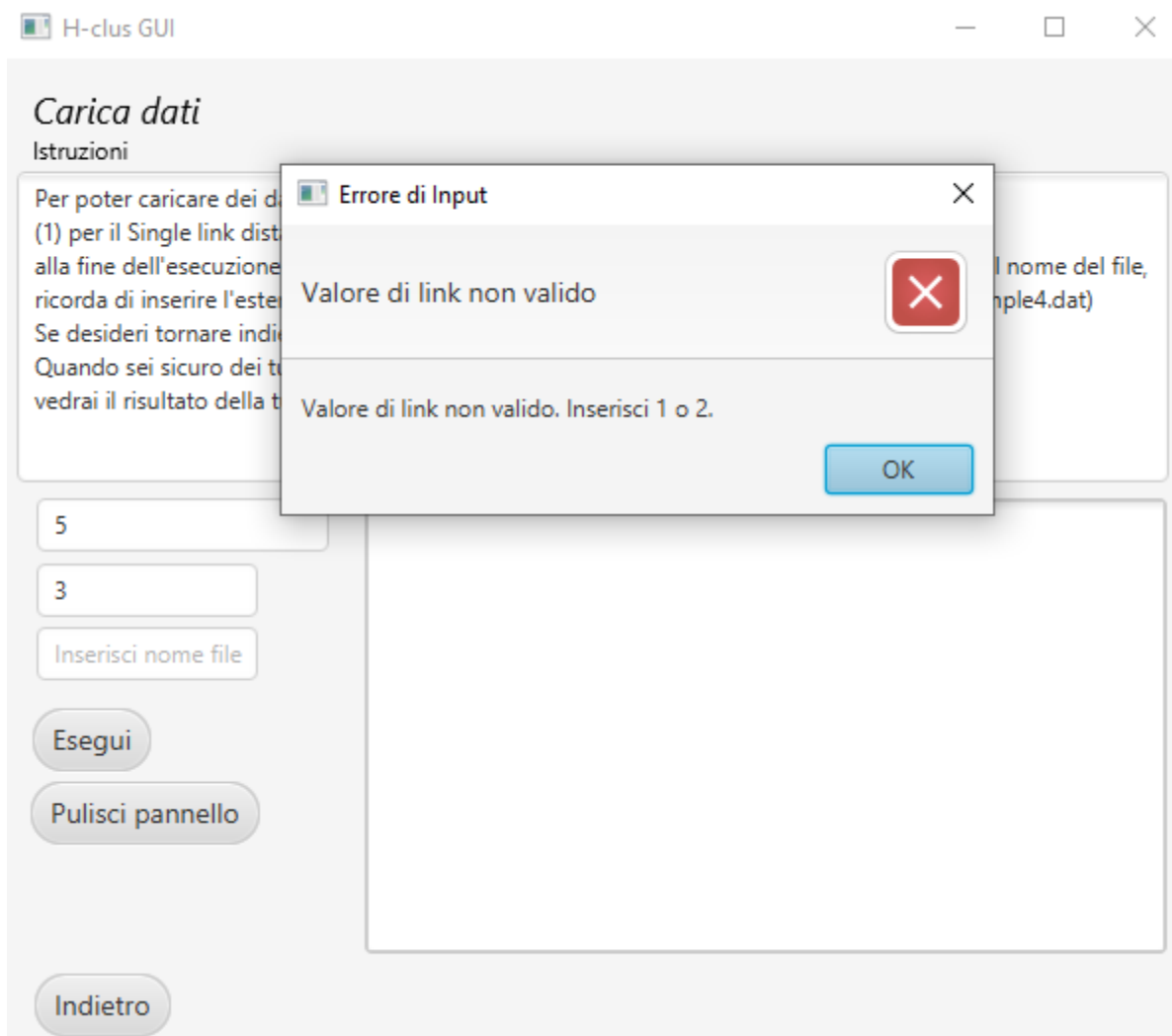
Cliccando su *“Indietro”*, si torna alla schermata iniziale. Cliccando su *“Pulisci pannello”* si svuota il pannello di output dei risultati. Cliccando su *“Esegui”* si inoltra la richiesta al server con gli inserimenti della profondità, il single o average link distance e il nome del file.

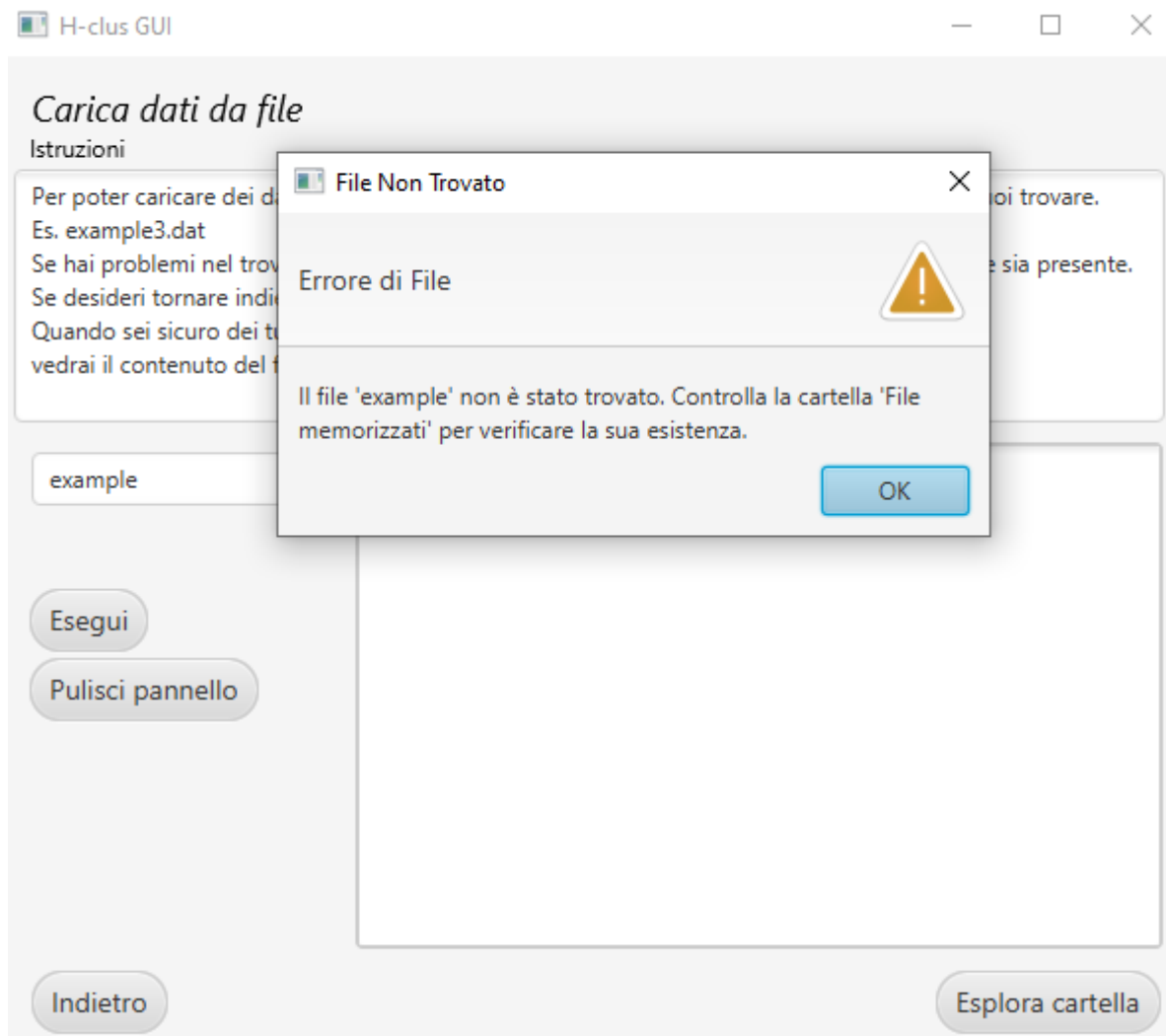
4) Carica dati da file:



Cliccando su *“Indietro”*, si torna alla schermata iniziale. Cliccando su *“Pulisci pannello”* si svuota il pannello di output dei risultati. Cliccando su *“Esplora cartella”* si apre la cartella *“File memorizzati”*. Cliccando su *“Esegui”* si inoltra la richiesta al server con l’inserimento del nome del file.

5) Casi particolari





Per uscire dal programma, premere il tasto chiudi in qualsiasi schermata oppure tornare alla schermata principale e premere il tasto *“Chiudi applicazione”*.

NOTE

ⁱ In alternativa si può aprire il file con un editor di testo e copiare il contenuto nella shell MySQL ⁱⁱ Nel caso in cui il file jar non funzionasse a causa di librerie non linkate correttamente, è possibile importare queste ultime come librerie esterne al progetto, ed avviare il client direttamente da IDE ⁱⁱⁱ Per passare dalla versione base a quella estesa o viceversa, assicurarsi di usare la giusta versione del server (\\Estesa\\). Se necessario, chiudere il server base prima di aprire il server esteso