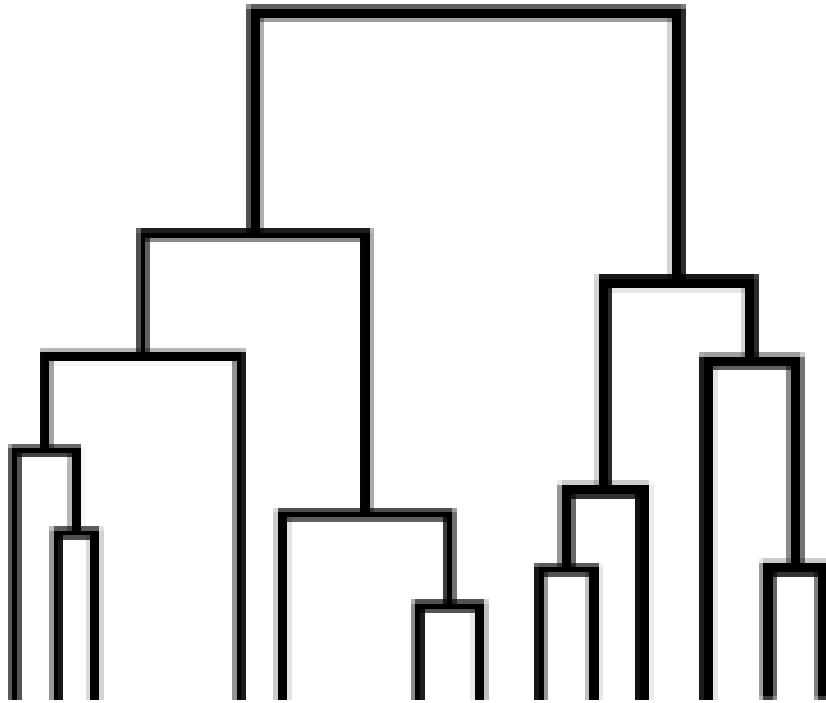


H-CLUS



*Emanuele Fornaro
Alessandro Boffolo Aldo
Vincenzo Pio Eraclea*

INDICE

Introduzione	2
Da chi è stato realizzato il progetto	4
Come è stato realizzato il progetto	4
Funzionalità implementate	5
Progetto base	5
Versione estesa	5
Guida utente	6
Prerequisiti	6
Versione base	6
Tipologie di errori	15

INTRODUZIONE

Il caso di studio analizzato nell'anno 2023-24 ha visto come soggetto la scoperta di un dendrogramma di cluster di dati con algoritmo di clustering agglomerativo.

Dati:

- Una collezione D di transazioni, dove ogni transazione è un vettore di valori misurati per una collezione di attributi numerici.
- Un intero k, che rappresenta il numero desiderato di cluster.

Scopo del Clustering:

Lo scopo del clustering è partizionare D in k insiemi di transazioni D_1, \dots, D_k , tale che:

- D_i ($i=1, \dots, k$) è un segmento (selezione) omogenea di D;
- $D = \cup_{i=1}^k D_i$, $D_i \cap D_j = \emptyset$

Calcolo delle Distanze:

Per calcolare la distanza tra due esempi si utilizza la distanza euclidea. Dati due cluster, possono essere calcolate due differenti distanze:

1. Distanza Single-link:

$$D(C1, C2) = \min_{(t1 \in C1, t2 \in C2)} dist(t1, t2)$$

2. Distanza Average-link:

$$D(C1, C2) = \frac{\sum_{t1 \in C1, t2 \in C2} d(t1, t2)}{\#(C1 \times C2)}$$

Clustering Agglomerativo:

Il clustering agglomerativo è un metodo di clustering gerarchico utilizzato nell'analisi dei dati e nell'apprendimento automatico per raggruppare insiemi di dati simili. Il clustering agglomerativo inizia considerando ogni punto come un cluster separato e procede unendo iterativamente i cluster più simili, fino a quando tutti i punti di dati appartengono a un unico cluster. Questo processo può essere dettagliato nei seguenti passaggi:

1. Inizializzazione:

- Ogni punto di dati è considerato un cluster separato.

2. Calcolo delle Distanze:

- Calcolare la matrice delle distanze tra tutti i punti utilizzando la distanza euclidea.

3. Unione dei Cluster:

- Identificare i due cluster più vicini utilizzando una delle metriche di distanza definite (single-link o average-link).
- Unire i due cluster selezionati.
- Aggiornare la matrice delle distanze per riflettere la fusione appena avvenuta.

4. Ripetizione:

- Ripetere il processo di unione dei cluster fino a quando rimane un solo cluster che contiene tutti i punti di dati.

Costruzione del Dendrogramma:

Durante il processo di fusione, viene costruito un dendrogramma, che è una rappresentazione grafica ad albero delle fusioni dei cluster. Il dendrogramma aiuta a visualizzare i vari livelli di similarità tra i cluster e può essere tagliato a diversi livelli per ottenere un numero desiderato di cluster.

Interpretazione del Dendrogramma:

Una volta costruito, il dendrogramma può essere analizzato per determinare il numero ottimale di cluster tagliandolo a un livello specifico. Il livello di taglio è scelto in modo tale da ottenere k cluster, dove k è il numero desiderato di cluster specificato inizialmente.

DA CHI È STATO REALIZZATO IL PROGETTO

- Boffolo Alessandro Aldo (mat.: 735963)
E-mail: a.boffolo@studenti.uniba.it
- Fornaro Emanuele (mat.:738721)
E-mail: e.fornaro4@studenti.uniba.it
- Eraclea Vincenzo Pio (mat.775561)
E-mail: v.eraclea@studenti.uniba.it

COME È STATO REALIZZATO IL PROGETTO

La struttura del progetto si basa sul linguaggio Java, inoltre è stato utilizzato *MySQL Workbench 8.0 CE* per poter leggere i data target della base di dati.

L'interfaccia utente è stata creata con *JavaFX* e con *SceneBuilder*. Essa semplifica l'utilizzo del programma, rendendolo più chiaro, grazie all'inserimento di elementi grafici e di finestre.

FUNZIONALITÀ IMPLEMENTATE

PROGETTO BASE

Nella versione base del programma è possibile leggere o scoprire i vari dendrogrammi da una base di dati, attraverso una scelta che viene effettuata sotto forma di menu testuale.

Per prima cosa si dovrà inserire il nome della tabella all'interno della base di dati da cui estrapolare i dati, scegliere se caricare da file o visualizzare da database il dendrogramma, inserire la profondità del dendrogramma di cluster e selezionare la distanza: se single-link o average-link.

Una volta visualizzati, è anche possibile salvare i dati in un file, dando prima il nome (compresa l'estensione), che potranno essere caricati in seguito con l'opzione di lettura.

Con l'opzione lettura è possibile visualizzare un file contenente il dendrogramma di cluster scoperto in precedenza, inserendo il nome di tale file (compresa l'estensione).

È stata implementata, inoltre, la gestione degli errori, attraverso la quale si cerca di spiegare nel modo più chiaro possibile all'utente eventuali problemi riscontrati durante la comunicazione tra client e server e/o trasmissioni con il database, fornendo una soluzione.

Essendo, inoltre, il server multithreading, è consentita una comunicazione tra un server e più client in tempo reale e con connessioni multiple grazie all'estensione della classe Thread di Java.

VERSIONE ESTESA

Nella versione estesa il programma offre la possibilità di scoprire o leggere un dendrogramma di cluster di dati con algoritmo agglomerativo da una base di dati attraverso l'uso di una interfaccia grafica.

All'avvio viene richiesto di inserire l'Indirizzo IP (127.0.0.1) e il Numero di Porta (8080), dopodiché si passa a una nuova schermata dove si potrà scegliere se scoprire o leggere il dendrogramma di cluster.

Nel caso si vogliano scoprire il dendrogramma di cluster si prosegue inserendo prima il nome della tabella all'interno della base di dati, in seguito la profondità del dendrogramma che si vuole scoprire e infine si dà la possibilità di scegliere una distanza tra single-link e average-link. Successivamente si potrà salvare in un file scrivendo il nome che si vuole dare al file (compresa l'estensione).

Nel caso si scelga invece l'opzione di lettura, bisognerà inserire il nome del file salvato (compresa l'estensione) contenente il dendrogramma di cluster già scoperto ed eseguirlo per poterlo visualizzare.

GUIDA UTENTE

PREREQUISITI

Java - Per l'esecuzione del programma (client o server, versione base o estesa) è necessario che sul computer sia stato installato Java, e che la versione dell'ambiente Java (JRE) sia uguale o superiore alla 1.8.

Per la JDK (Java Development Kit) è stata utilizzata la versione 21.

Link di download Java: <https://www.java.com/it/download/>

Ulteriori informazioni: <https://www.java.com/it/download/help/path.html>

VERSIONE BASE

H-Clus si compone di due programmi: un programma client e un programma server. Per poter permettere la comunicazione, è necessario che l'applicativo server sia avviato, precedentemente rispetto all'applicativo client, su un qualsiasi host.

Preconfigurazione

1) Avvio MultiServer (**Server**)

Per avviare *MultiServer* bisognerà aggiungere alla propria configurazione di Run un numero di porta valido su cui ospitare il servizio Server (con valori al di fuori del range 1-1024 e con un massimo di 65535).

Nella configurazione fornita il numero di porta inserito sarà 8080, se la si vuole cambiare bisognerà seguire i seguenti passi:

selezionare **"Edit configurations..."**, nella sezione **"Program arguments"** inserire un valore valido, per esempio **8080** -> premere su **"Apply"** per applicare le modifiche.

In caso di "Errore di connessione al database: com.mysql.cj.jdbc.Driver" assicurarsi di aver collegato il Driver di MySQL, eseguire i seguenti passi:

```
File -> Project Structure -> Modules -> Server U-CLUS->
Dependencies -> + -> JARs or Directories... ->
navigate\to\the\driver\path -> premere su "mysql-connector-java-
8.0.17.jar" ->Apply
```

Infine, assicurarsi di avere eseguito lo Script MySQL *createDB* sulla macchina che ospita il Server, in modo da caricare all'interno del DataBase la tabella.

2) Avvio MainTest (**Client**)

Per avviare *MainTest* aggiungere alla propria configurazione di Run un IP address e un numero di porta su cui il Server è in ascolto. La configurazione fornita presenta come indirizzo IP localhost "127.0.0.1" e il numero di porta "8080".

Per cambiare la configurazione bisognerà seguire i seguenti passi:

"Edit configurations...", nella sezione "Program arguments" inserire l'indirizzo IP e il numero di porta su cui il Server fornisce il servizio es. -> 127.0.0.1 8080 -> premere "Apply"

3) Avvio Estensione (Extension)

In caso di "Errore JavaFX runtime components are missing..." assicurarsi di avere javafx-sdk collegato (incluso nel file del progetto). Seguire i seguenti passi:

1. creare una libreria JavaFX in Idea

```
File -> Project Structure -> Libraries -> + -> Point to the  
lib folder of the JavaFX SDK -> Apply
```

2. Aggiungere le opzioni VM per il module path

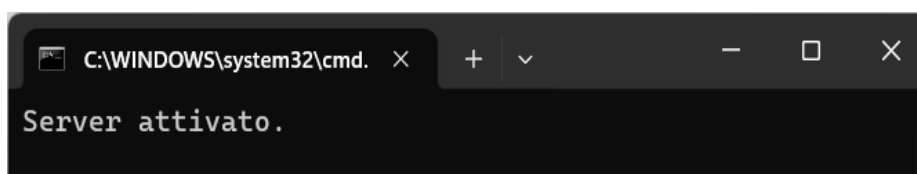
```
Run -> Edit Configurations -> Modify options -> Add VM  
options -> --module-path  
out\artifacts\Extension_jar\javafx-sdk-21.0.1\lib --add-  
modules  
javafx.controls,javafx.fxml,javafx.graphics,javafx.base -  
> Apply
```

Avvio dell'applicativo (versione Base)

Una volta configurato, se si vuole aprire il programma, bisognerà aprire le seguenti cartelle:

ServerH-CLUS -> out -> artifacts -> ServerH-CLUS_jar

Ed eseguire prima il file *ServerH-CLUS.bat*, che mostrerà questa schermata:



In seguito, per avviare la versione base, bisognerà avviare il file *ClientH-CLUS.bat*, attraverso il percorso:

ClientH-CLUS -> out -> artifacts -> ClientH-CLUS_jar

Nel menu che comparirà bisognerà inserire il nome della tabella all'interno della base di dati che si vuole utilizzare.


```
C:\WINDOWS\system32\cmd. x + v
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=49439]
Nome tabella:
exampleTab
```

Apprendi Dendrogramma da Database

Se si vuole scoprire il dendrogramma di cluster bisognerà digitare 2, la profondità del dendrogramma desiderato e quale distanza utilizzare, digitando 1 per usare **single-link distance** e 2 per usare **average-link distance**.

```
C:\WINDOWS\system32\cmd. x + v
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=49439]
Nome tabella:
exampleTab
Scegli una opzione
(1) Carica Dendrogramma da File
(2) Apprendi Dendrogramma da Database
Risposta:2
Introdurre la profondita' del dendrogramma
3
Distanza: single-link (1), average-link (2):
1
```

Una volta visualizzati è anche possibile salvare i dati in un file, attribuendogli un nome, inserendo anche l'estensione che si desidera utilizzare (come nell'esempio qui sotto dove il file è stato chiamato *prova.txt*).

```
C:\WINDOWS\system32\cmd. x + v
3
Distanza: single-link (1), average-link (2):
1
level0:
cluster0:<[1.0,2.0,0.0]>
cluster1:<[0.0,1.0,-1.0]>
cluster2:<[1.0,3.0,5.0]>
cluster3:<[1.0,3.0,4.0]>
cluster4:<[2.0,2.0,0.0]>

level1:
cluster0:<[1.0,2.0,0.0]><[2.0,2.0,0.0]>
cluster1:<[0.0,1.0,-1.0]>
cluster2:<[1.0,3.0,5.0]>
cluster3:<[1.0,3.0,4.0]>

level2:
cluster0:<[1.0,2.0,0.0]><[2.0,2.0,0.0]>
cluster1:<[0.0,1.0,-1.0]>
cluster2:<[1.0,3.0,5.0]><[1.0,3.0,4.0]>

Inserire il nome dell'archivio (comprensivo di estensione):
prova.txt
```

Carica Dendrogramma da File

Invece, digitando 1 nel menu di scelta, è possibile visualizzare un file contenente il dendrogramma di cluster scoperto in precedenza, inserendo il nome di tale file, compreso di estensione (es. *prova.txt*).

```
C:\WINDOWS\system32\cmd. x + v
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=49449]
Nome tabella:
exampleTab
Scegli una opzione
(1) Carica Dendrogramma da File
(2) Apprendi Dendrogramma da Database
Risposta:1
Inserire il nome dell'archivio (comprensivo di estensione):
prova.txt
level0:
cluster0:<[1.0,2.0,0.0]>
cluster1:<[0.0,1.0,-1.0]>
cluster2:<[1.0,3.0,5.0]>
cluster3:<[1.0,3.0,4.0]>
cluster4:<[2.0,2.0,0.0]>

level1:
cluster0:<[1.0,2.0,0.0]><[2.0,2.0,0.0]>
cluster1:<[0.0,1.0,-1.0]>
cluster2:<[1.0,3.0,5.0]>
cluster3:<[1.0,3.0,4.0]>

level2:
cluster0:<[1.0,2.0,0.0]><[2.0,2.0,0.0]>
cluster1:<[0.0,1.0,-1.0]>
cluster2:<[1.0,3.0,5.0]><[1.0,3.0,4.0]>

Premere un tasto per continuare . . .
```

Avvio dell'applicativo (versione Estesa)

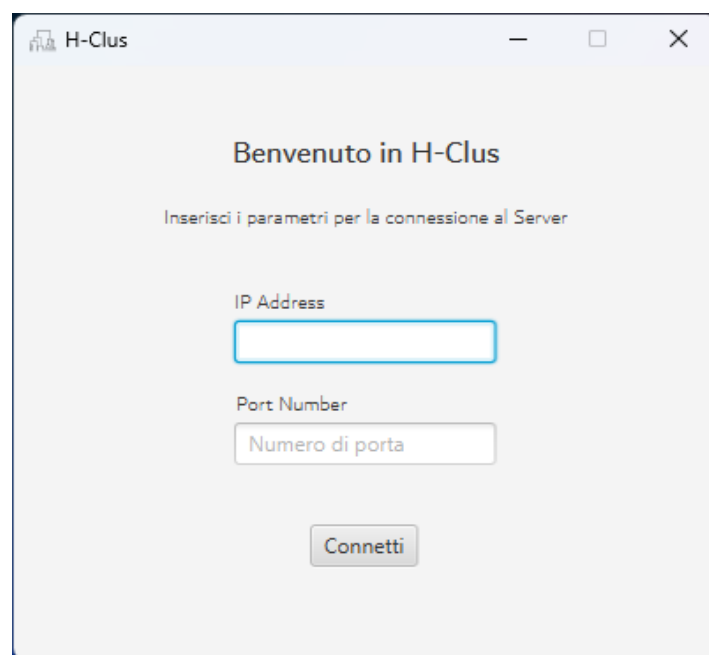
La versione estesa presenta le stesse funzionalità del programma base, ma con l'implementazione di un'interfaccia grafica, il cui scopo è quello di rendere l'applicativo user-friendly. Se la si vuole avviare bisognerà eseguire il file *ServerH-CLUS.bat* attraverso il percorso:

ServerH-CLUS -> out -> artifacts -> ServerH-CLUS_jar

Ed in seguito eseguire il file *Extension.bat* attraverso il percorso:

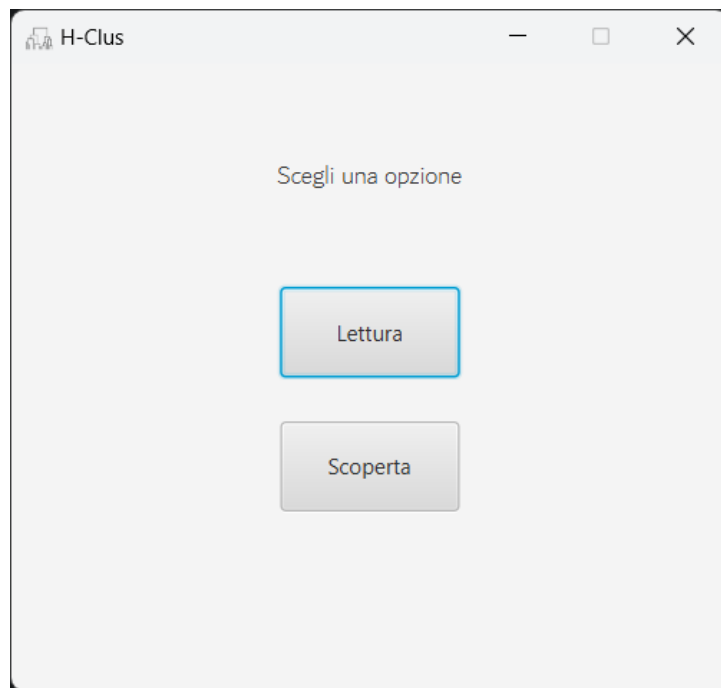
Extension -> Extension -> out -> artifacts -> Extension_jar

che mostrerà la seguente schermata:

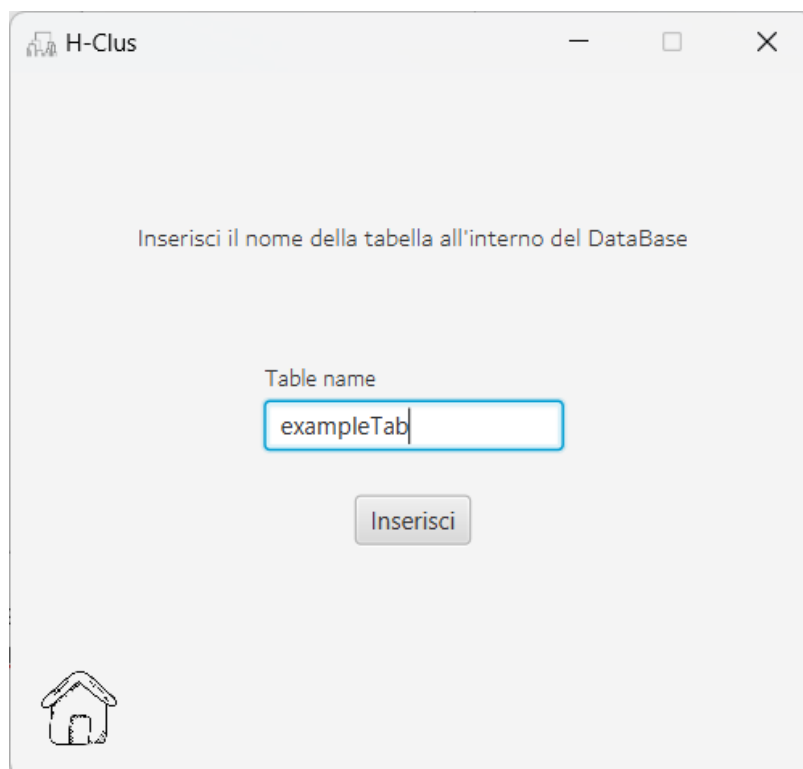
The image shows a screenshot of a Java application window titled "H-Clus". The window has a light gray background and a standard Windows-style title bar with minimize, maximize, and close buttons. The main content area displays the text "Benvenuto in H-Clus" in a bold, black font. Below this, a smaller line of text reads "Inserisci i parametri per la connessione al Server". There are two input fields: the first is labeled "IP Address" and the second is labeled "Port Number" with the placeholder text "Numero di porta" inside it. Both fields have a light blue border. At the bottom center of the form is a button labeled "Connetti".

Qui viene richiesto di inserire l'Indirizzo IP (es. **127.0.0.1**) e il Numero di Porta (es. **8080**). Una volta inseriti bisognerà premere sul pulsante **Connetti**.

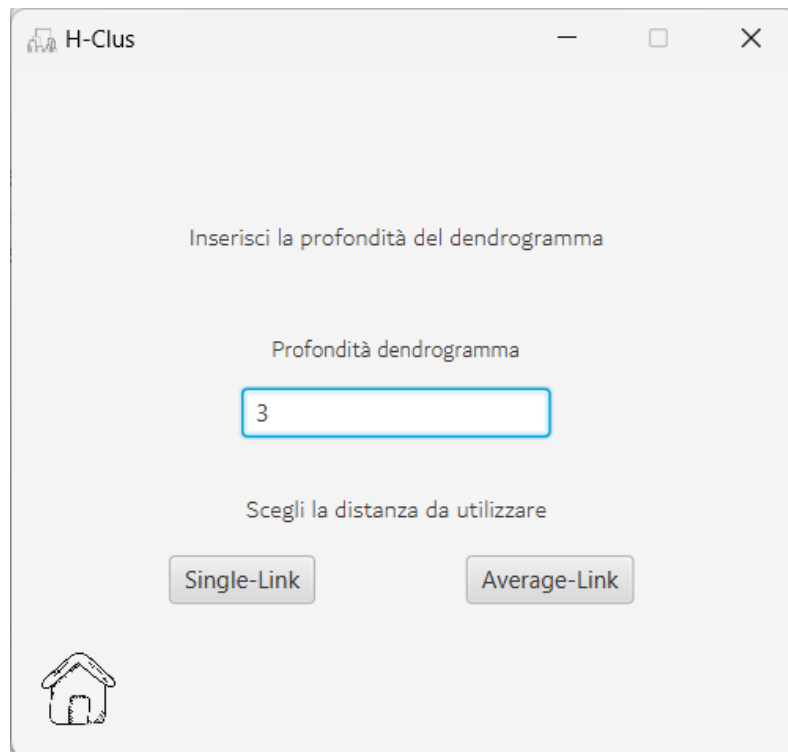
Dopodiché, se l'indirizzo IP e il Numero di Porta sono validi, visualizzeremo una nuova schermata (*Schermata principale*) dove si potrà scegliere se scoprire o leggere il dendrogramma di cluster.



Premendo il pulsante **Scoperta** verrà visualizzata la seguente schermata dove sarà richiesto di inserire il nome della tabella all'interno della base di dati dalla quale scoprire il dendrogramma di cluster.

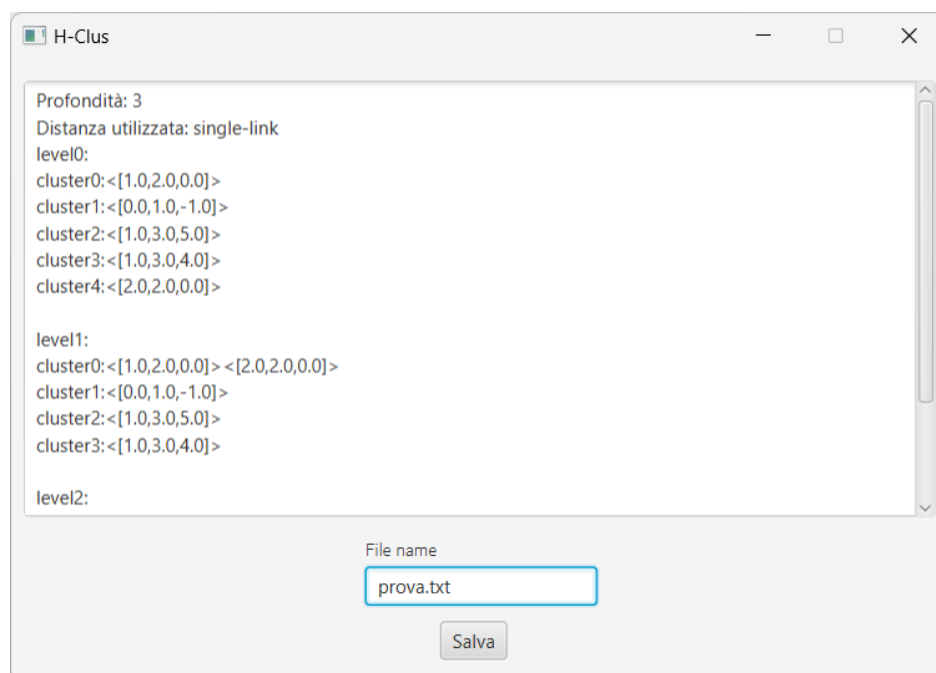


Premendo il pulsante **Inserisci** si passa a una nuova schermata dove verrà richiesto prima di inserire la profondità del dendrogramma che si preferisce visualizzare nella parte inferiore della schermata, poi, in basso, si potrà scegliere il tipo di distanza da utilizzare: **Single-Link** o **Average-Link**.



The screenshot shows the H-Clus application window. At the top, the title bar says "H-Clus". The main area contains the text "Inserisci la profondità del dendrogramma". Below this is a text input field labeled "Profondità dendrogramma" with the number "3" entered. Underneath is the text "Scegli la distanza da utilizzare". There are two buttons: "Single-Link" and "Average-Link". At the bottom left, there is a small house icon.

Dopo aver inserito correttamente i dati e aver premuto su uno dei due pulsanti verrà aperta una nuova finestra contenente il dendrogramma di cluster scoperti, secondo la tipologia di distanza che si è preferito utilizzare, che dovranno essere salvati in un file, inserendo il nome (compresa l'estensione), come nell'immagine sottostante, e premendo il pulsante **Salva**.



The screenshot shows the H-Clus application window displaying the results of a clustering process. The title bar says "H-Clus". The main area contains the following text:

```
Profondità: 3
Distanza utilizzata: single-link
level0:
cluster0: <[1.0,2.0,0.0]>
cluster1: <[0.0,1.0,-1.0]>
cluster2: <[1.0,3.0,5.0]>
cluster3: <[1.0,3.0,4.0]>
cluster4: <[2.0,2.0,0.0]>

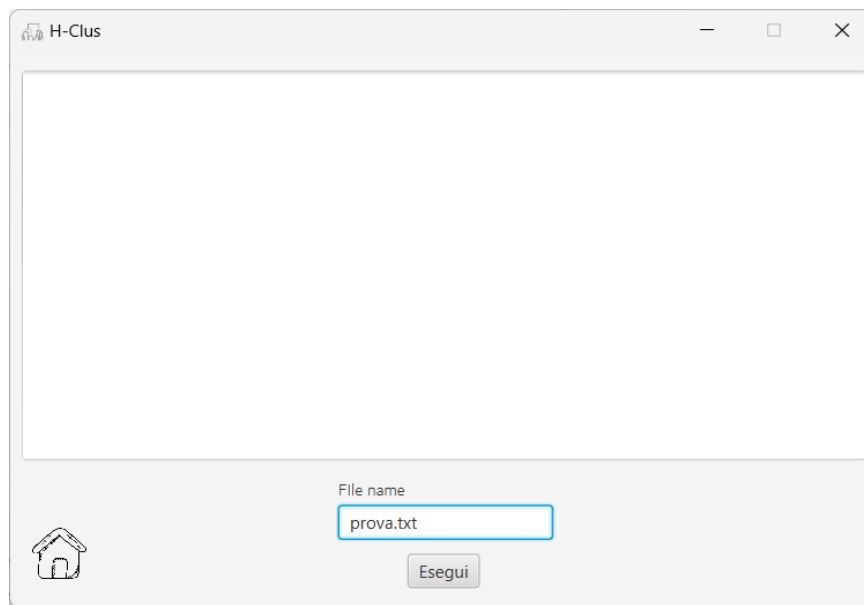
level1:
cluster0: <[1.0,2.0,0.0]> <[2.0,2.0,0.0]>
cluster1: <[0.0,1.0,-1.0]>
cluster2: <[1.0,3.0,5.0]>
cluster3: <[1.0,3.0,4.0]>

level2:
```

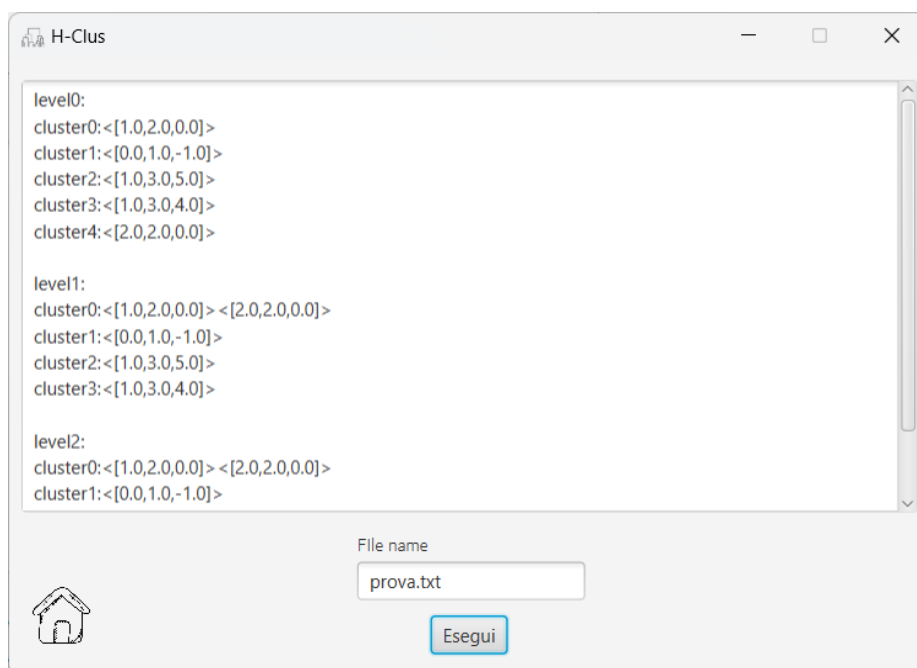
At the bottom, there is a text input field labeled "File name" with the text "prova.txt" entered. Below this is a button labeled "Salva".

Infine, basterà chiudere questa finestra per poter tornare a quella precedente, dove si potrà inserire un nuovo valore per il dendrogramma di cluster da visualizzare all'interno della stessa tabella, oppure si potrà tornare alla schermata principale per poter scegliere se leggere o scoprire il dendrogramma, premendo sul pulsante a forma di casetta in basso a destra.

Premendo il pulsante **Lettura** nella Schermata Principale sarà possibile visualizzare un file contenente il dendrogramma di cluster scoperti in precedenza, inserendo il nome di tale file (compresa l'estensione) e premendo il pulsante **Esegui** come nella schermata seguente.



Nella schermata successiva viene visualizzato il file.

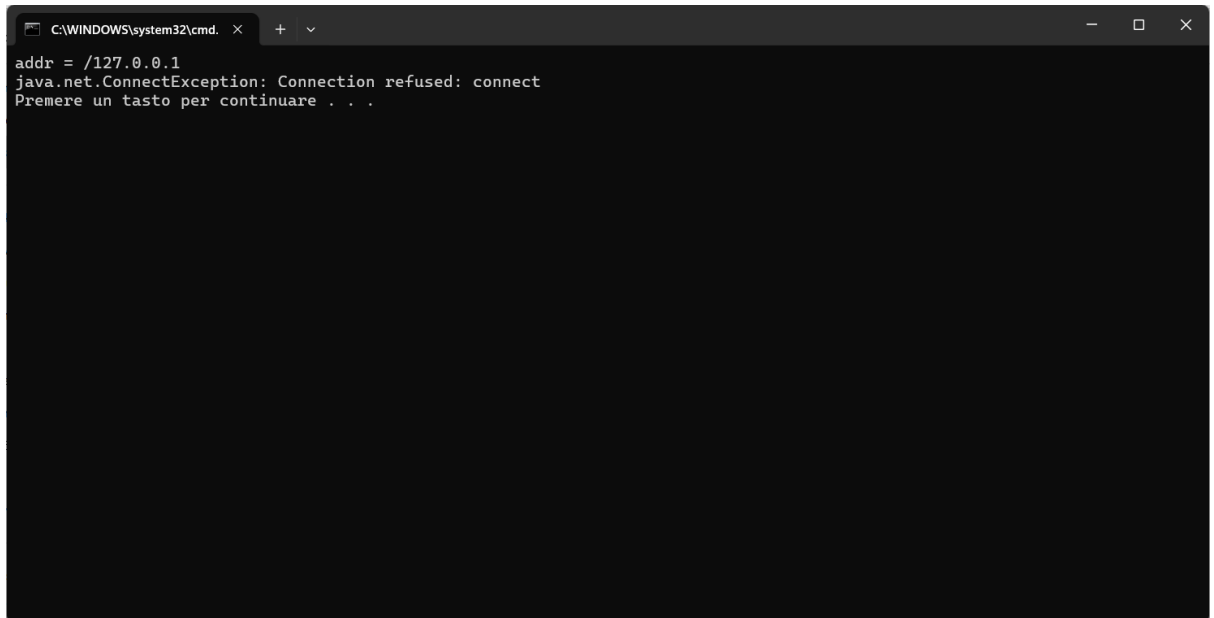


Premendo sul pulsante a forma di casetta si tornerà alla *Schermata Principale*.

TIPOLOGIE DI ERRORI

Errore server non attivo

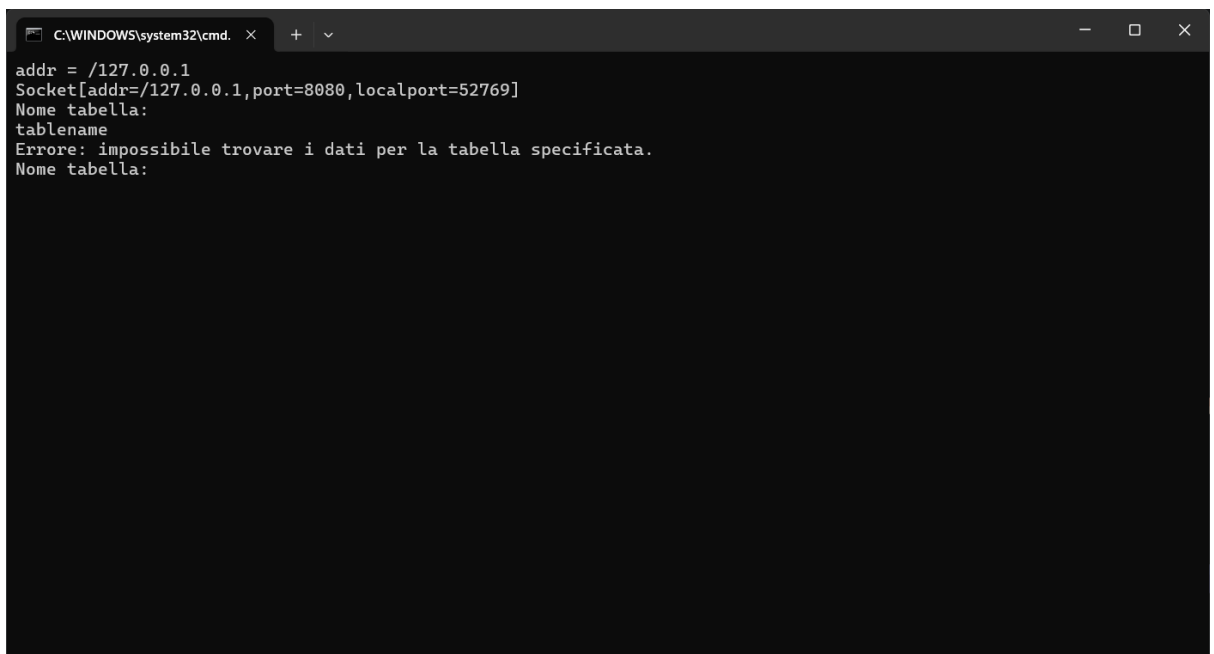
Aprendo prima il file Client del file Server verrà visualizzato il seguente messaggio.



```
C:\WINDOWS\system32\cmd. x + v
addr = /127.0.0.1
java.net.ConnectException: Connection refused: connect
Premere un tasto per continuare . . .
```

Errore nome della tabella del database

Inserendo nel menù un nome di una tabella di database non presente verrà visualizzato il seguente messaggio.



```
C:\WINDOWS\system32\cmd. x + v
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=52769]
Nome tabella:
tablename
Errore: impossibile trovare i dati per la tabella specificata.
Nome tabella:
```

Errore profondità del dendrogramma di cluster da scoprire

Inserendo un valore numerico per la profondità del dendrogramma di cluster da voler scoprire maggiore o minore di quello consentito verrà visualizzato il seguente messaggio.

```
C:\WINDOWS\system32\cmd. x + v
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=49472]
Nome tabella:
exampleTab
Scegli una opzione
(1) Carica Dendrogramma da File
(2) Apprendi Dendrogramma da Database
Risposta:2
Introdurre la profondita' del dendrogramma
0
Distanza: single-link (1), average-link (2):
1
Errore: la profondita' specificata non è valida. Fornire un valore corretto.
Premere un tasto per continuare . . .
```

Errore profondità del dendrogramma di cluster da scoprire inserendo una lettera

Inserendo un valore non numerico per la profondità del dendrogramma di cluster da voler scoprire verrà visualizzato il seguente messaggio.

```
C:\WINDOWS\system32\cmd. x + v
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=49478]
Nome tabella:
exampleTab
Scegli una opzione
(1) Carica Dendrogramma da File
(2) Apprendi Dendrogramma da Database
Risposta:2
Introdurre la profondita' del dendrogramma
a
Errore nella lettura dei dati di tipo int, è stato restituito il valore MIN_VALUE.
Distanza: single-link (1), average-link (2):
1
Errore: la profondita' specificata non è valida. Fornire un valore corretto.
Premere un tasto per continuare . . .
```

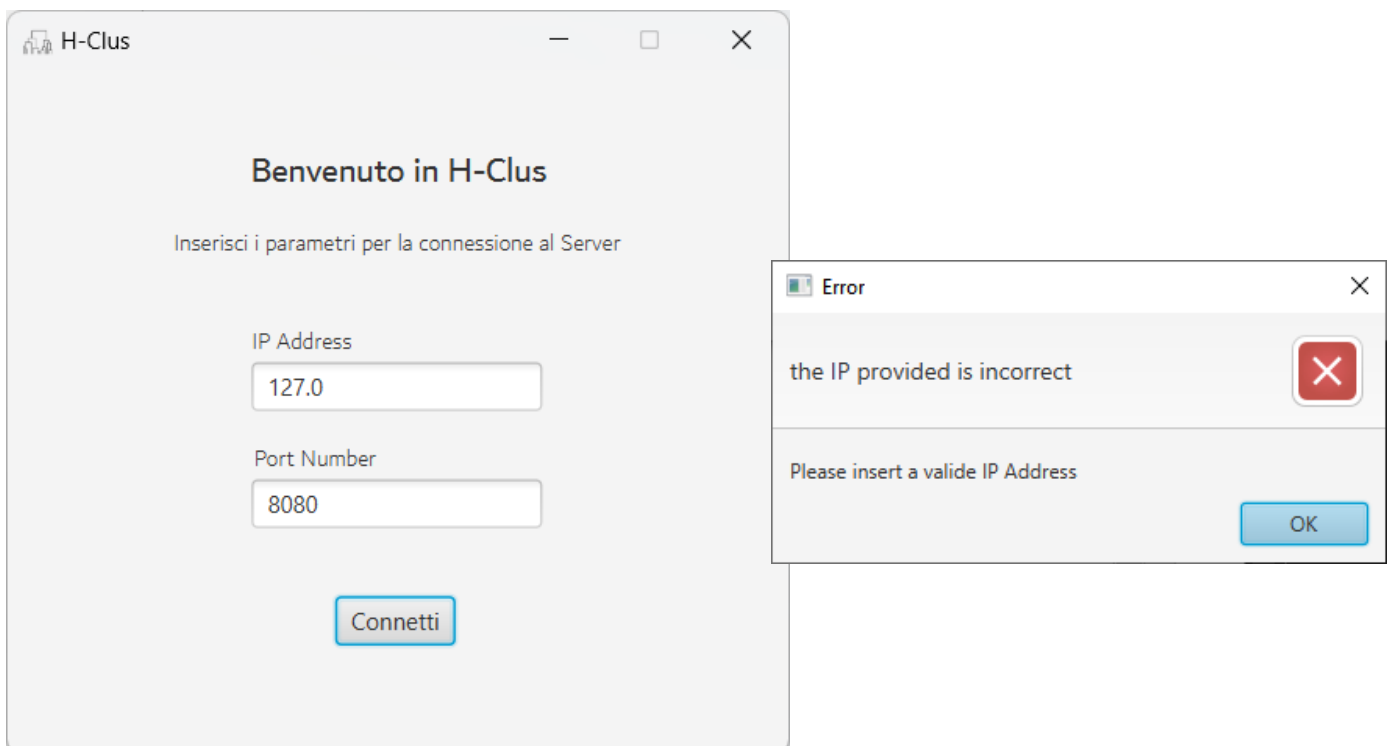

Errore nome del file in cui è serializzato il dendrogramma di cluster da recuperare

Inserendo il nome di un file non presente tra quelli salvati dalla quale visualizzare il dendrogramma di cluster scoperti verrà visualizzato il seguente messaggio.

```
C:\WINDOWS\system32\cmd. x + v
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=52769]
Nome tabella:
tablename
Errore: impossibile trovare i dati per la tabella specificata.
Nome tabella:
exampleTab
Scegli una opzione
(1) Carica Dendrogramma da File
(2) Apprendi Dendrogramma da Database
Risposta:1
Inserire il nome dell'archivio (comprensivo di estensione):
filename.txt
Errore: file non trovato. Controlla il percorso del file.
Premere un tasto per continuare . . .
```

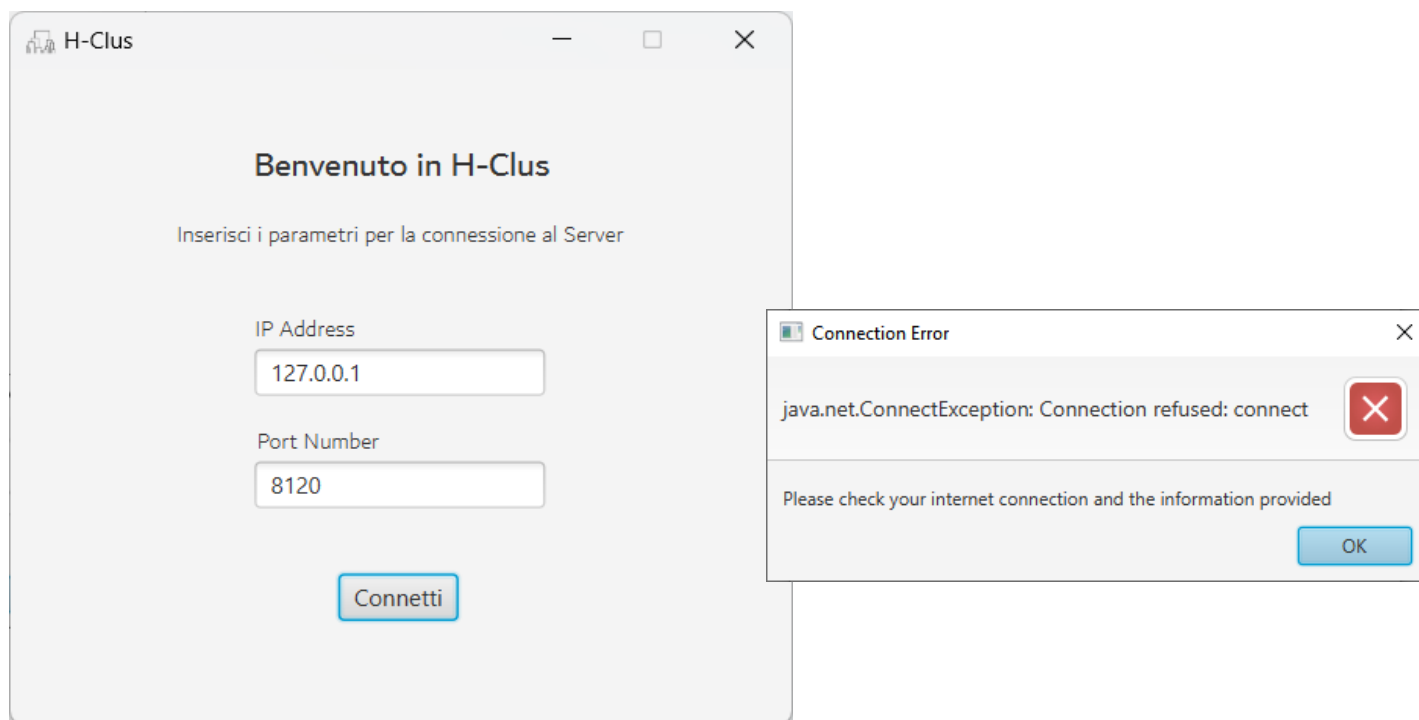
Errore Indirizzo IP non valido (Versione Estesa)

Inserendo un Indirizzo IP non valido verrà visualizzato il seguente messaggio.



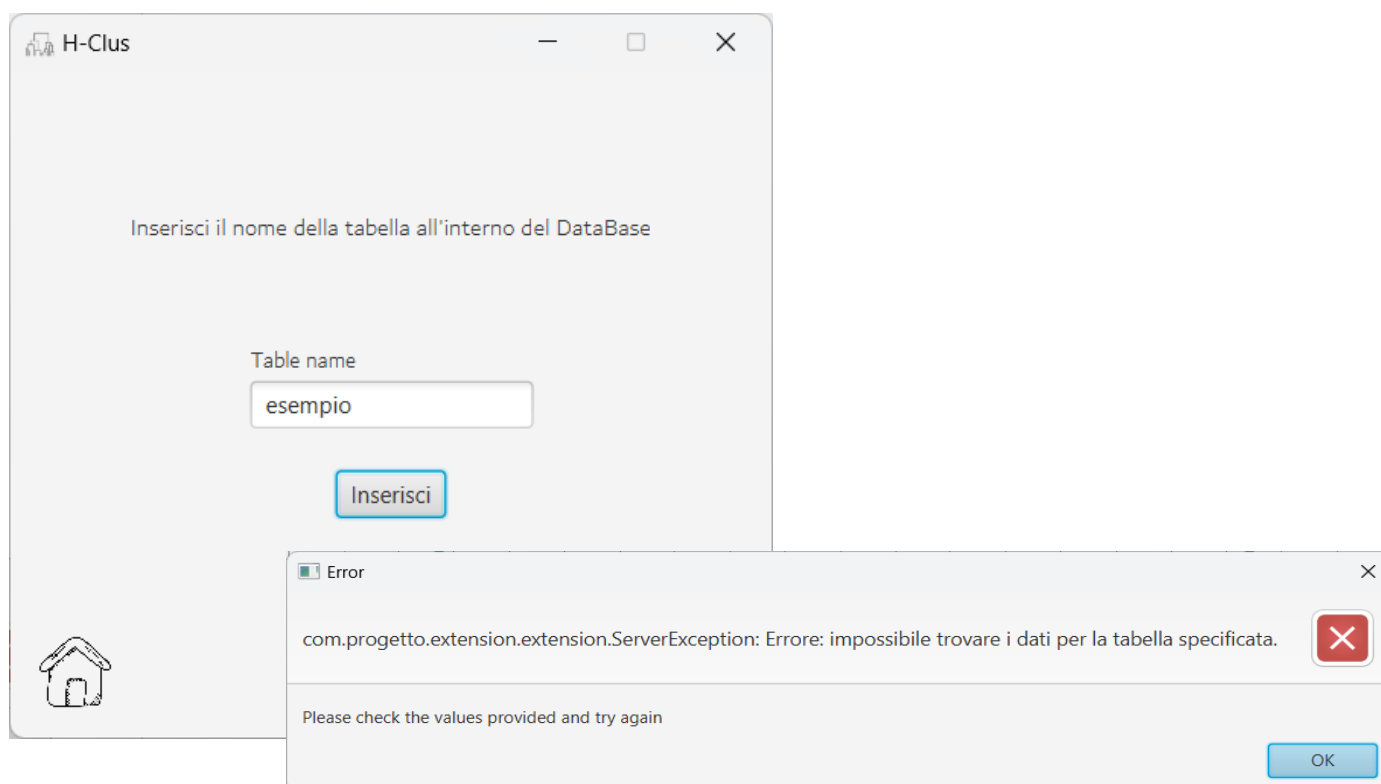
Errore Numero di Porta (versione Estesa)

Inserendo un Numero di Porta non valido verrà visualizzato il seguente messaggio.



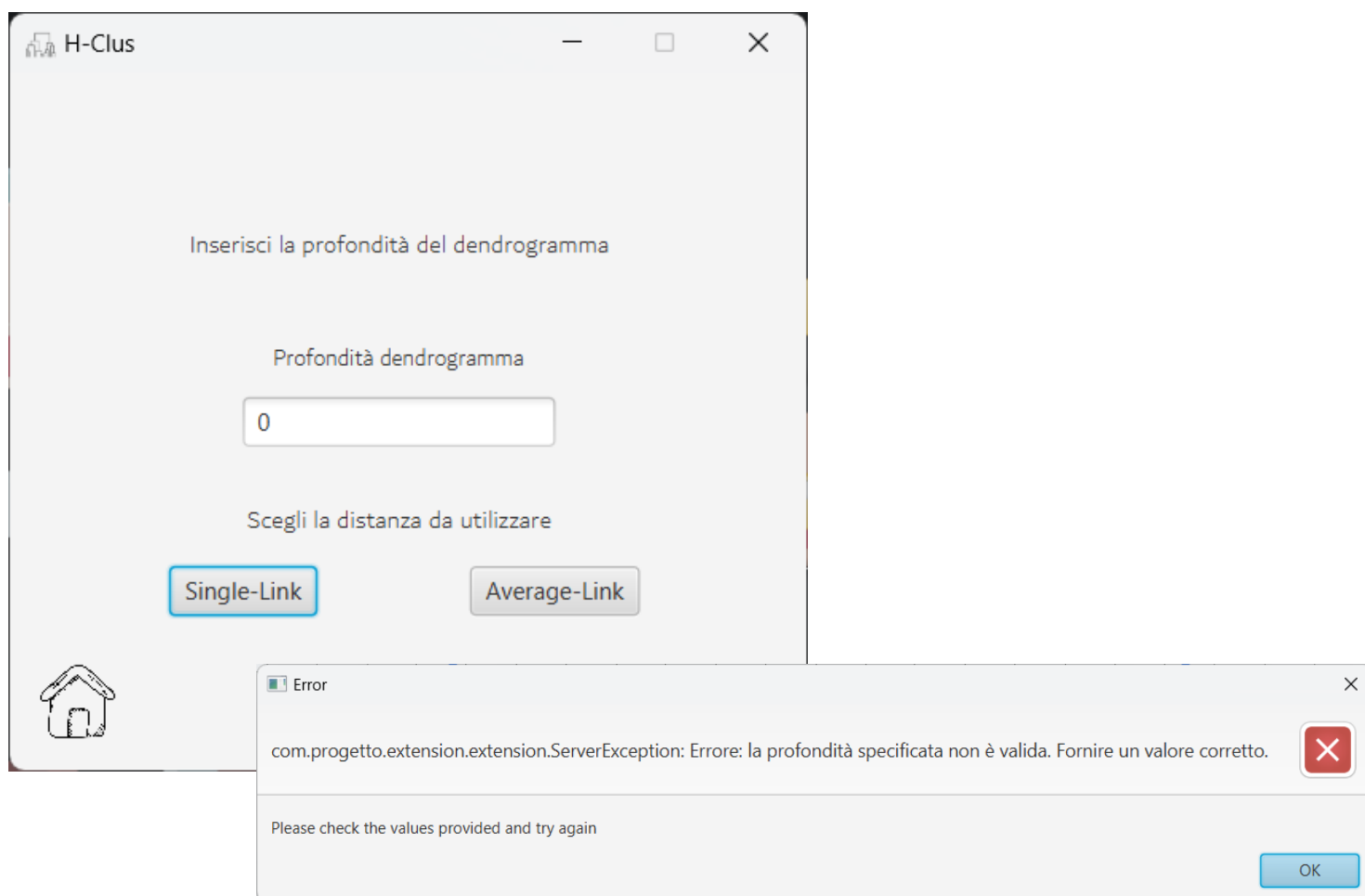
Errore nome della tabella di database (versione Estesa)

Inserendo nel menù un nome di una tabella di database non presente verrà visualizzato il seguente messaggio.



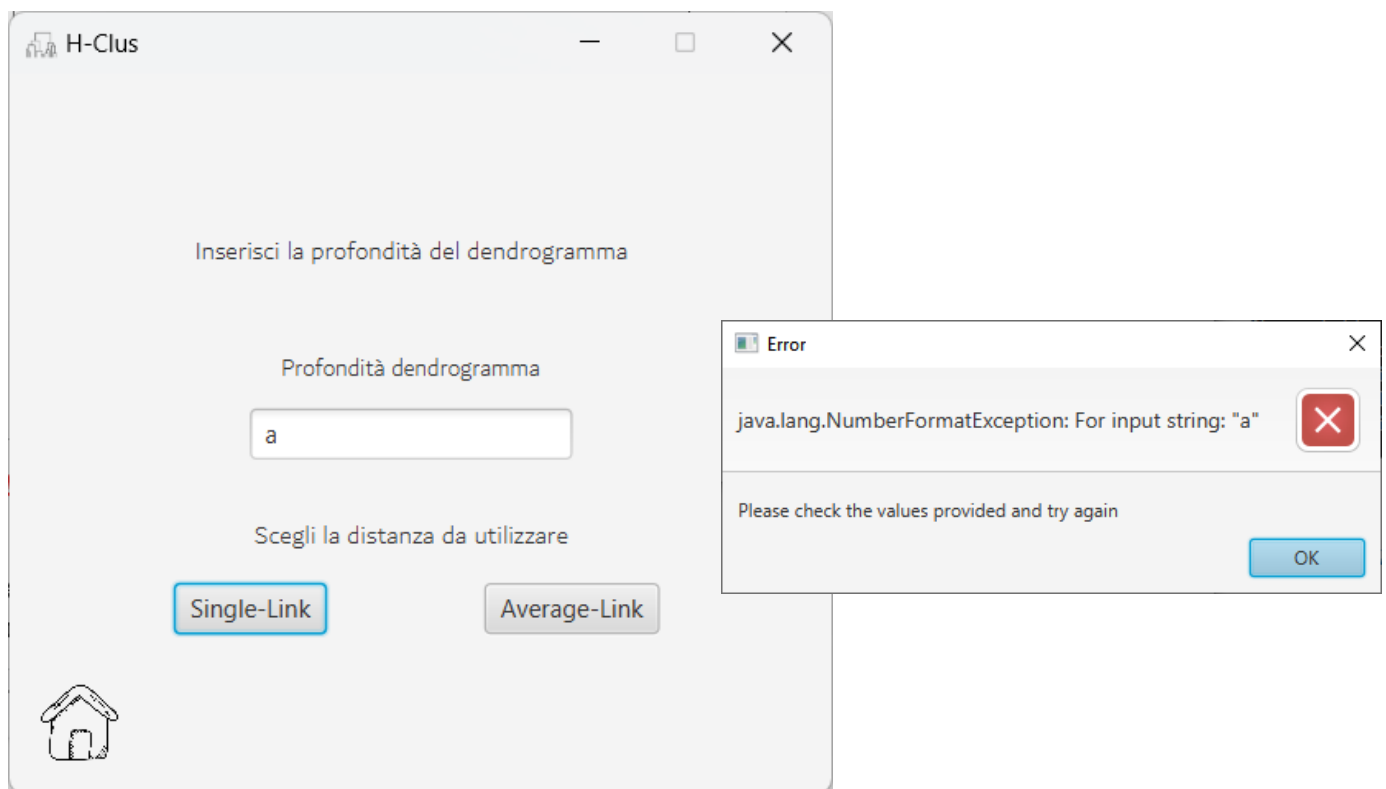
Errore profondità del dendrogramma di cluster da scoprire (versione Estesa)

Inserendo un valore numerico per la profondità del dendrogramma di cluster da voler scoprire maggiore o minore di quello consentito verrà visualizzato il seguente messaggio.



Errore profondità del dendrogramma di cluster da scoprire inserendo una lettera (versione Estesa)

Inserendo un valore non numerico per la profondità del dendrogramma di cluster da voler scoprire verrà visualizzato il seguente messaggio.



Errore nome del file in cui è serializzato il dendrogramma di cluster da recuperare (versione Estesa)

Inserendo il nome di un file non presente tra quelli salvati dalla quale visualizzare il dendrogramma di cluster scoperti verrà visualizzato il seguente messaggio.

