

DOCUMENTAZIONE

Progetto realizzato da:

- Jacopo D'Abramo (mat. 716484)
mail: j.dabramo@studenti.uniba.it
- Lorenzo Cassano (mat. 718331)
mail: l.cassano25@studenti.uniba.it
- Stefano Bavaro (mat. 716034)
mail: s.bavaro4@studenti.uniba.it

Il software presentato permette, a partire da un dataset memorizzato tramite il servizio MySQL, la ricerca di *pattern frequenti* e l'analisi di *pattern emergenti*, sulla base di un valore minimo di *supporto* e un valore di *grow rate* forniti dall'utente. La tecnica di *data mining* che viene implementata sfrutta un [algoritmo apriori](#) fornito dal docente.

Definiamo un **pattern frequente** come un sottoinsieme di *Item*, ove un *item* è una coppia [attributo, valore], che occorrono con una frequenza minima rappresentata dal *minimo supporto*, in un dato insieme.

Un **pattern emergente** è definito come un insieme di item la cui frequenza cambia in maniera notevole da un dataset ad un altro (*grow rate*, tasso di crescita).

Progetto base

La versione base del progetto consiste in un'architettura client/server che permette all'utente di esplorare un dataset creato dal server.

Il server dovrà essere eseguito su una macchina con un database MySQL in esecuzione.

Il servizio sarà raggiungibile sulla porta 8080, e potrà comunicare con diversi client contemporaneamente.

I servizi offerti dal server sono i seguenti:

- Caricamento del Database contenente il *dataset*
- Ricerca di *frequent pattern* dato un *minimo supporto*
- Scoperta di *emerging pattern* sulla base di un valore di *grow rate* e *minimo supporto*
- Salvataggio e caricamento su file di *frequent* ed *emerging pattern*

Il client, da riga di comando, permette di collegarsi ad una macchina che sta eseguendo un'istanza del server (ServerSocket), il cui indirizzo IP e porta devono essere specificati

all'avvio del client.

L'utente dovrà, successivamente, specificare se effettuare una nuova ricerca, oppure consultare l'archivio per reperire frequent pattern ed emerging pattern già trovati in precedenza. In entrambi i casi l'utente sarà tenuto ad inserire i nomi della tabella target e quella di background insieme ai valori di minimo supporto e di grow rate.

Una volta effettuata la scelta, si avvierà l'esplorazione del dataset o dell'archivio, tramite **l'algoritmo APRIORI**. Alla fine del processo, all'utente sarà presentato l'output della ricerca. Sarà quindi possibile ricominciare una nuova ricerca o interrogare l'archivio. Il salvataggio dell'output verrà effettuato in automatico in seguito a una nuova ricerca. L'utente sarà informato di eventuali errori sollevati dal server durante l'esecuzione delle operazioni richieste.

Progetto Esteso

L'estensione per il progetto consiste in un'interfaccia grafica per desktop, sviluppata esclusivamente lato client (il server non richiede interazione con l'utente). Tale interfaccia è stata sviluppata tramite l'applicativo **Scene Builder** per quanto concerne la progettazione grafica delle finestre e quindi dei relativi *bottoni*, *campi di testo* e *label*. Mentre per quanto riguarda la gestione del codice il tool utilizzato è stato il framework **javaFX**, ovvero una libreria Java che contiene una serie di classi e metodi per la gestione delle finestre menzionate precedentemente.

L'estensione prevede un'interfaccia grafica suddivisa in 3 finestre principali:

- La prima finestra viene mostrata all'avvio dell'applicativo e consente all'utente di avere a disposizione un *Help command* e di potersi connettere al server tramite un indirizzo e una port.
- La seconda finestra viene mostrata se la connessione con il server non ha prodotto errori. Essa permette all'utente di inserire tutti i parametri necessari per poter eseguire una ricerca nel dataset.
- La terza e ultima finestra si occupa di mostrare a video l'output dell'elaborazione del server.

Guida di installazione

I file *.jar* *.bat* sono riposti nella cartella *bin* di ogni progetto; vi è anche una cartella *bin* per l'intero progetto contenente due file *.bat*, che permettono di avviare rispettivamente il server e client base oppure il server e client esteso.

Installazione Server

Per eseguire il Server sulla propria macchina è necessario:

- Aver installato la *Java RunTime Enviroment* 8 o versioni successive
- Aver installato *MySQL*
- Eseguire il Servizio *MySQL*

Installazione Java

Per poter installare la *Java RunTime Enviroment* basta visitare il seguente link:

<https://www.java.com/it/download/manual.jsp> .

Nel caso fosse già installata la *Java RunTime* può essere verificata la sua versione immettendo il comando `java -version` sulla *Comand Line*.

Installazione MySQL

MySQL è il *Database Management System* usato nel progetto e può essere scaricato dal seguente link <https://dev.mysql.com/downloads/installer/>.

Una volta installato *MySQL* esso dovrà essere eseguito per poter avviare l'applicazione.

Per poter avviare il servizio *MySQL* basta accedere all'elenco dei *Servizi*, trovare il servizio *MySQL* ed eseguirlo.

Installazione Client

Per poter eseguire correttamente il Client (sia del progetto base che del progetto esteso) bisogna:

- Aver installato la *Java RunTime Enviroment* 8 o versioni successive
- Avere un server in ascolto

NB: Il Client del progetto esteso potrà essere avviato anche senza nessun Server in ascolto ma al momento in cui ci si dovrà collegare al Server (prima finestra di apertura del Client del progetto esteso) verranno sollevati degli errori di connessione.

Guida Utente

Avvio del Server

Per avviare il server è necessario eseguire il file *avvio server.bat* nella cartella *Server/bin*. Il server può anche essere eseguito da riga di comando, tramite l'istruzione

```
java -jar Server.jar
```

Avvio del Client (progetto base)

Per avviare il client è necessario eseguire il file *avvio client base.bat* nella cartella *ClientBase/bin*; in questo modo il client si conatterà al server in esecuzione sulla propria macchina sulla porta 8080. Il client può anche essere eseguito da riga di comando, tramite l'istruzione

```
java -jar client_base.jar [INDIRIZZO] [PORTA]
```

ovvero

```
java -jar client_base.jar localhost 8080
```

Avvio del Client (progetto esteso)

Per avviare il client del progetto esteso è necessario eseguire il file *avvio client_esteso.bat* nella cartella *Client/bin*; in questo modo verrà aperta una finestra dove dovranno essere inserite l'indirizzo e la porta del server sul quale ci si vuole connettere. Il client può anche essere eseguito da riga di comando, tramite l'istruzione

```
java -jar Client_esteso.jar
```

Avvio del Server e del Client

All'interno della cartella *bin* si trovano due file *.bat* :

- *avvio_server_e_client_base.bat* il quale avvia automaticamente il server e il client del progetto base facendolo collegare automaticamente al server.
- *avvio_server_e_client_esteso* il quale avvia automaticamente il server e il client del progetto esteso con l'apertura della prima finestra per il collegamento al server.

Guida di utilizzo

Utilizzo Client (progetto base)

Una volta avviato il client del progetto base e la connessione è andata a buon fine si specifica quale operazione si vuole effettuare:

- nuova scoperta (digitando 1)
- risultati in archivio (digitando 2)

```
addr = localhost/127.0.0.1
port=8080
Socket[addr=localhost/127.0.0.1,port=8080,localport=56826]
Scegli una opzione:
1:Nuova scoperta
2: Risultati in archivio
```

Una volta effettuata la scelta bisogna inserire i seguenti dati:

- valore minimo di supporto
- growrate (o tasso di crescita)
- tabella target
- tabella di background

Questi dati saranno inviati al server.

Successivamente ci sarà la ricerca dei pattern nel dataset in base ai criteri inseriti che verranno inviati al client e stampati a video.

```
Scegli una opzione:
1:Nuova scoperta
2: Risultati in archivio
1
Inserire valore minimo supporto (minsup>0 e minsup<=1):
0.3
Inserire valore minimo grow rate (minGr>0):
1
Tabella target:
palytennistarget
Tabella background:
playtennisbackground
Errore di connessione al Database
Vuoi ripetere?(s/n)
s
Scegli una opzione:
1:Nuova scoperta
2: Risultati in archivio
1
Inserire valore minimo supporto (minsup>0 e minsup<=1):
0.3
Inserire valore minimo grow rate (minGr>0):
1
Tabella target:
playtennistarget
Tabella background:
playtennisbackground
```

```
1: (outlook=rain)[0.375]
2: (outlook=sunny)[0.375]
3: (temperature in [0.0;6.062])[0.375]
4: (temperature in [24.248;30.31])[0.375]
5: (umidity=normal)[0.375]
6: (wind=strong)[0.375]
7: (outlook=sunny) AND (umidity=high)[0.375]
8: (outlook=sunny) AND (play=no)[0.375]
9: (temperature in [0.0;6.062]) AND (umidity=normal)[0.375]
10: (temperature in [24.248;30.31]) AND (umidity=high)[0.375]
11: (umidity=high) AND (outlook=sunny)[0.375]
12: (umidity=high) AND (temperature in [24.248;30.31])[0.375]
13: (umidity=high) AND (play=no)[0.375]
14: (umidity=normal) AND (temperature in [0.0;6.062])[0.375]
15: (wind=weak) AND (play=yes)[0.375]
16: (play=no) AND (outlook=sunny)[0.375]
17: (play=no) AND (umidity=high)[0.375]
18: (play=yes) AND (wind=weak)[0.375]
19: (outlook=sunny) AND (umidity=high) AND (play=no)[0.375]
20: (outlook=sunny) AND (play=no) AND (umidity=high)[0.375]
21: (umidity=high) AND (outlook=sunny) AND (play=no)[0.375]
22: (umidity=high) AND (play=no) AND (outlook=sunny)[0.375]
23: (play=no) AND (outlook=sunny) AND (umidity=high)[0.375]
24: (play=no) AND (umidity=high) AND (outlook=sunny)[0.375]
25: (play=no)[0.5]
26: (play=yes)[0.5]
27: (umidity=high) AND (wind=weak)[0.5]
28: (wind=weak) AND (umidity=high)[0.5]
```

```
29: (umidity=high)[0.625]
30: (wind=weak)[0.625]

1: (outlook=rain)[0.375] [1.125]
2: (outlook=sunny)[0.375] [1.125]
3: (wind=weak)[0.625] [1.25]
4: (umidity=high)[0.625] [1.875]
5: (temperature in [0.0;6.062])[0.375] [2.25]
6: (temperature in [24.248;30.31])[0.375] [2.25]
7: (temperature in [0.0;6.062]) AND (umidity=normal)[0.375] [2.25]
8: (umidity=high) AND (play=no)[0.375] [2.25]
9: (umidity=normal) AND (temperature in [0.0;6.062])[0.375] [2.25]
10: (play=no) AND (umidity=high)[0.375] [2.25]
11: (play=no)[0.5] [3.0]
12: (outlook=sunny) AND (umidity=high)[0.375] [Infinity]
13: (outlook=sunny) AND (play=no)[0.375] [Infinity]
14: (temperature in [24.248;30.31]) AND (umidity=high)[0.375] [Infinity]
15: (umidity=high) AND (outlook=sunny)[0.375] [Infinity]
16: (umidity=high) AND (temperature in [24.248;30.31])[0.375] [Infinity]
17: (play=no) AND (outlook=sunny)[0.375] [Infinity]
18: (outlook=sunny) AND (umidity=high) AND (play=no)[0.375] [Infinity]
19: (outlook=sunny) AND (play=no) AND (umidity=high)[0.375] [Infinity]
20: (umidity=high) AND (outlook=sunny) AND (play=no)[0.375] [Infinity]
21: (umidity=high) AND (play=no) AND (outlook=sunny)[0.375] [Infinity]
22: (play=no) AND (outlook=sunny) AND (umidity=high)[0.375] [Infinity]
23: (play=no) AND (umidity=high) AND (outlook=sunny)[0.375] [Infinity]
24: (umidity=high) AND (wind=weak)[0.5] [Infinity]
25: (wind=weak) AND (umidity=high)[0.5] [Infinity]
```

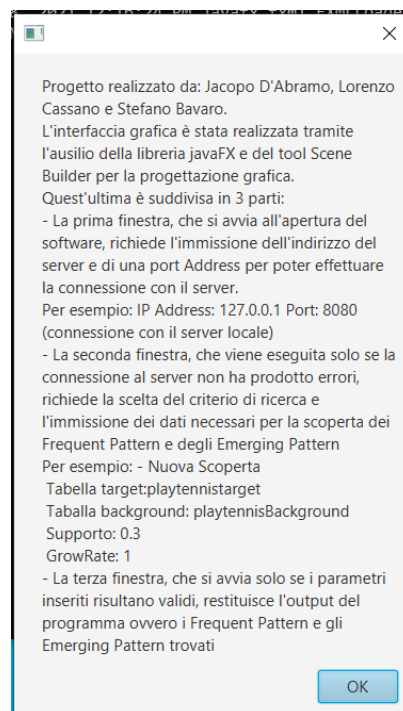
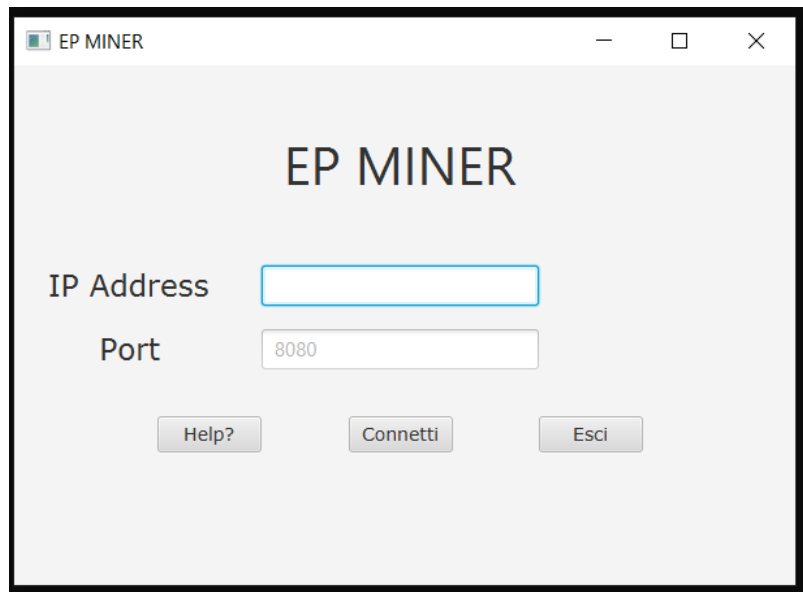
Utilizzo del Client (progetto esteso)

Schermata di connessione con il Server

All'avvio del client del progetto esteso verrà aperta una finestra dove dovranno essere inseriti indirizzo e porta del server per avviare la connessione.

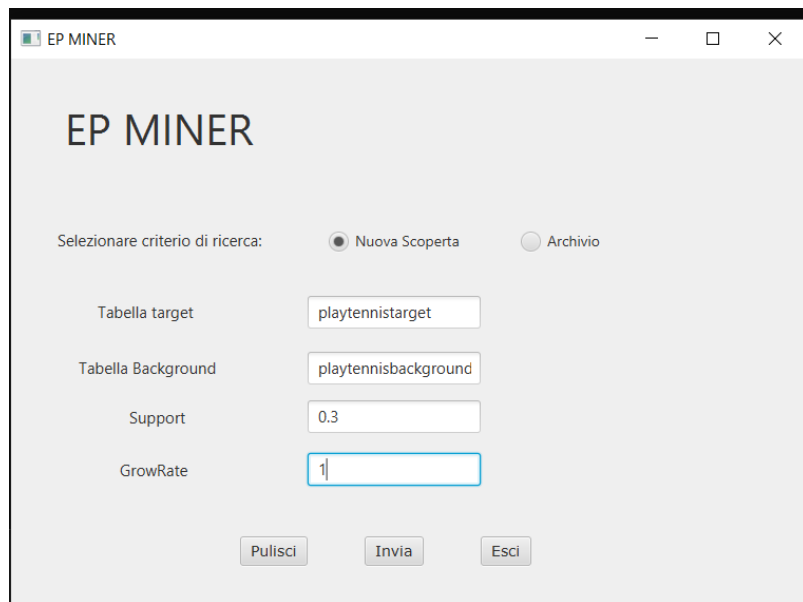
Nella finestra avremo tre bottoni:

- **Connetti** il quale conatterà il client con il server se i dati sono corretti.
- **Help** il quale una volta premuto stampa un avviso contenente le principali informazioni dell'applicazione.
- **Esci** il quale permette di uscire dall'applicazione.



Schermata di Inserimento dei dati

Appena il client si collega con il server si aprirà una finestra relativa all'inserimento dei dati (i dati da inserire sono gli stessi del client base) per ricercare i pattern nel dataset.



EP MINER

Selezionare criterio di ricerca: ☒ Nuova Scoperta ☐ Archivio

Tabella target: playtennistarget

Tabella Background: playtennisbackground

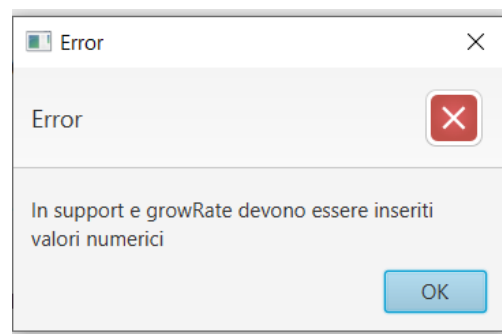
Support: 0.3

GrowRate: 1

Pulisci Invia Esci

Nel caso i dati non siano corretti verranno stampati a video degli avvisi relativi agli errori di inserimento dei dati.

Es



Nella schermata è anche presente un bottone pulisci il quale "pulisce" i campi dove inserire i dati per la ricerca dei pattern

Schermata di stampa dei risultati

Se non ci dovessero essere errori sull'inserimento dei dati il server effettuerà le ricerche dei pattern; se il server trova i pattern, il client aprirà una nuova finestra dove sono mostrati i risultati, altrimenti sarà mostrato a video un messaggio che avvisa che non sono stati trovati i pattern.

EP MINER

RISULTATI

Tabella Target: playtennistarget

Supporto: 0.3

Tabella Background: playtennisbackground

GrowRate: 1.0

Frequent Pattern
1: (outlook=rain)[0.375]
2: (outlook=sunny)[0.375]
3: (temperature in [0.0;6.062])[0.375]
4: (temperature in [24.248;30.31])[0.375]
5: (umidity=normal)[0.375]
6: (wind=strong)[0.375]
7: (outlook=sunny) AND (umidity=high)[0.375]
8: (outlook=sunny) AND (play=no)[0.375]
9: (temperature in [0.0;6.062]) AND (umidity=normal)[0.375]
10: (temperature in [24.248;30.31]) AND (umidity=high)[0.375]
11: (umidity=high) AND (outlook=sunny)[0.375]
12: (umidity=high) AND (temperature in [24.248;30.31])[0.375]
13: (umidity=high) AND (play=no)[0.375]
14: (umidity=normal) AND (temperature in [0.0;6.062])[0.375]
15: (wind=weak) AND (play=yes)[0.375]
16: (play=no) AND (outlook=sunny)[0.375]

Nuova Ricerca

Esci

EP MINER

RISULTATI

Tabella Target: playtennistarget

Supporto: 0.3

Tabella Background: playtennisbackground

GrowRate: 1.0

2: (outlook=sunny)[0.375] [1.125]
3: (wind=weak)[0.625] [1.25]
4: (umidity=high)[0.625] [1.875]
5: (temperature in [0.0;6.062])[0.375] [2.25]
6: (temperature in [24.248;30.31])[0.375] [2.25]
7: (temperature in [0.0;6.062]) AND (umidity=normal)[0.375] [2.25]
8: (umidity=high) AND (play=no)[0.375] [2.25]
9: (umidity=normal) AND (temperature in [0.0;6.062])[0.375] [2.25]
10: (play=no) AND (umidity=high)[0.375] [2.25]
11: (play=no)[0.5] [3.0]
12: (outlook=sunny) AND (umidity=high)[0.375] [Infinity]
13: (outlook=sunny) AND (play=no)[0.375] [Infinity]
14: (temperature in [24.248;30.31]) AND (umidity=high)[0.375] [Infinity]
15: (umidity=high) AND (outlook=sunny)[0.375] [Infinity]
16: (umidity=high) AND (temperature in [24.248;30.31])[0.375] [Infinity]
17: (play=no) AND (outlook=sunny)[0.375] [Infinity]
18: (outlook=sunny) AND (umidity=high) AND (play=no)[0.375] [Infinity]

Nuova Ricerca

Esci

Esempi di Test

Gli output numerici attesi sono stati ricavati dalla documentazione fornita insieme alle esercitazioni effettuate durante il corso.

Esempi di test validi

Input:

- tabella target = *"playtennistarget"*
- tabella backgorund = *"playtennisBackground"*
- minimo supporto = 0.3
- growrate = 1

Output:

EP MINER

RISULTATI

Tabella Target:playtennistarget Supporto:0.3

Tabella Background:playtennisBackground GrowRate:1.0

Frequent Pattern

- 1: (outlook=rain)[0.375]
- 2: (outlook=sunny)[0.375]
- 3: (temperature in [0.0;6.062])[0.375]
- 4: (temperature in [24.248;30.31])[0.375]
- 5: (umidity=normal)[0.375]
- 6: (wind=strong)[0.375]
- 7: (outlook=sunny) AND (umidity=high)[0.375]
- 8: (outlook=sunny) AND (play=no)[0.375]
- 9: (temperature in [0.0;6.062]) AND (umidity=normal)[0.375]
- 10: (temperature in [24.248;30.31]) AND (umidity=high)[0.375]
- 11: (umidity=high) AND (outlook=sunny)[0.375]
- 12: (umidity=high) AND (temperature in [24.248;30.31])[0.375]
- 13: (umidity=high) AND (play=no)[0.375]
- 14: (umidity=normal) AND (temperature in [0.0;6.062])[0.375]
- 15: (wind=weak) AND (play=yes)[0.375]
- 16: (play=no) AND (outlook=sunny)[0.375]
- 17: (play=no) AND (umidity=high)[0.375]

Nuova Ricerca Esci

EP MINER

RISULTATI

Tabella Target:playtennistarget Supporto:0.3

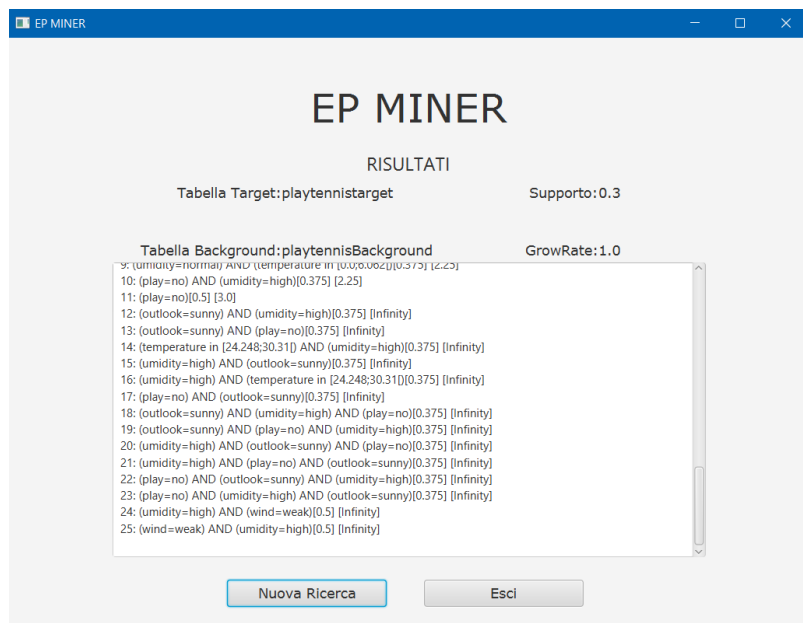
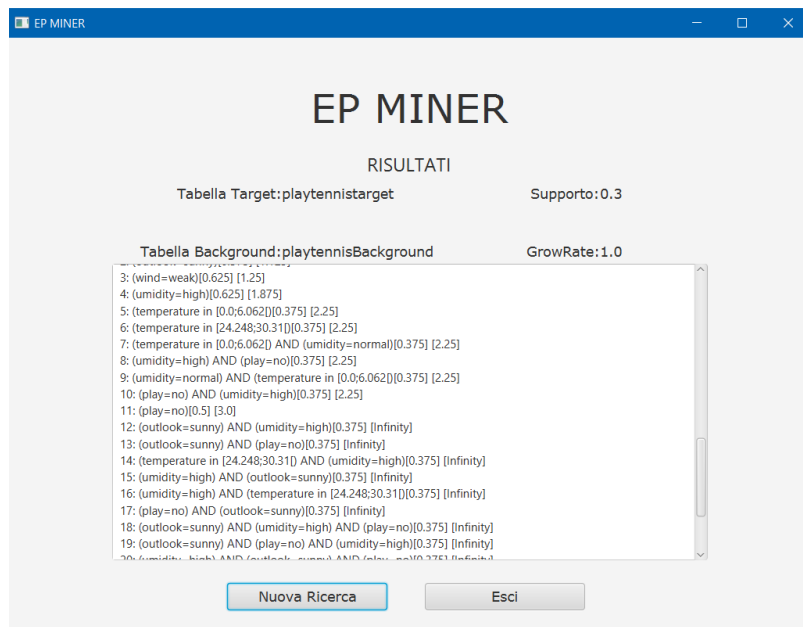
Tabella Background:playtennisBackground GrowRate:1.0

- 17: (play=no) AND (umidity=high)[0.375]
- 18: (play=yes) AND (wind=weak)[0.375]
- 19: (outlook=sunny) AND (umidity=high) AND (play=no)[0.375]
- 20: (outlook=sunny) AND (play=no) AND (umidity=high)[0.375]
- 21: (umidity=high) AND (outlook=sunny) AND (play=no)[0.375]
- 22: (umidity=high) AND (play=no) AND (outlook=sunny)[0.375]
- 23: (play=no) AND (outlook=sunny) AND (umidity=high)[0.375]
- 24: (play=no) AND (umidity=high) AND (outlook=sunny)[0.375]
- 25: (play=no)[0.5]
- 26: (play=yes)[0.5]
- 27: (umidity=high) AND (wind=weak)[0.5]
- 28: (wind=weak) AND (umidity=high)[0.5]
- 29: (umidity=high)[0.625]
- 30: (wind=weak)[0.625]

Emerging Pattern

- 1: (outlook=rain)[0.375] [1.125]
- 2: (outlook=sunny)[0.375] [1.125]

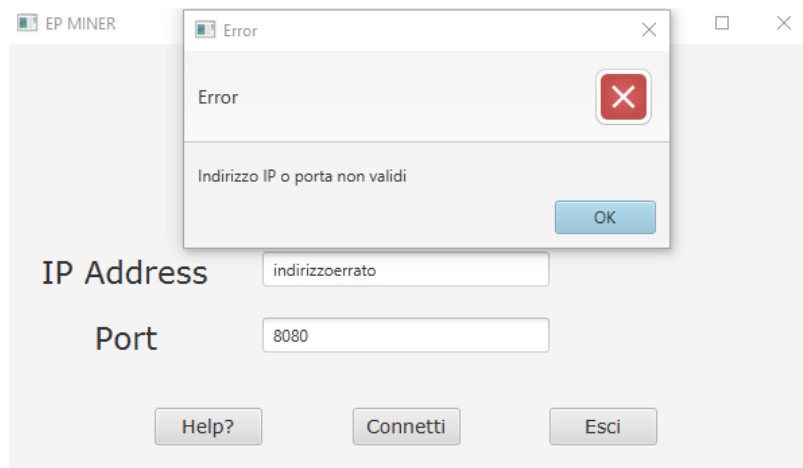
Nuova Ricerca Esci



Esempi di test contenenti errori

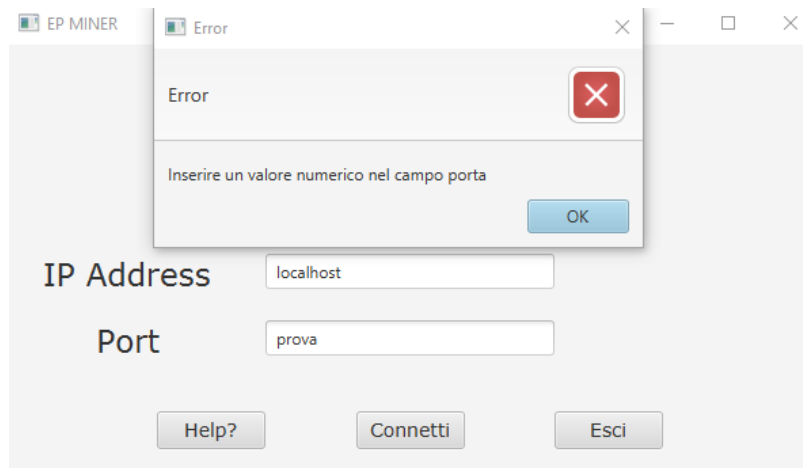
Input: indirizzo sbagliato

Output: messaggio di errore relativo all'indirizzo errato



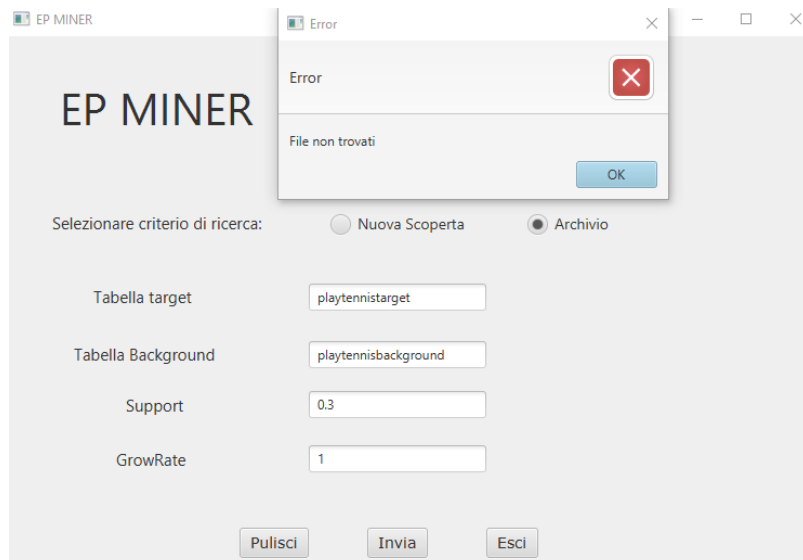
Input: valore numerico errato nel campo porta per il collegamento al server

Output: messaggio di errore relativo all'errato inserimento di un valore numerico nel campo porta per collegarsi al server



Input: dati relativi ad un file non esistente

Output: messaggio di errore relativo alla non presenza del file



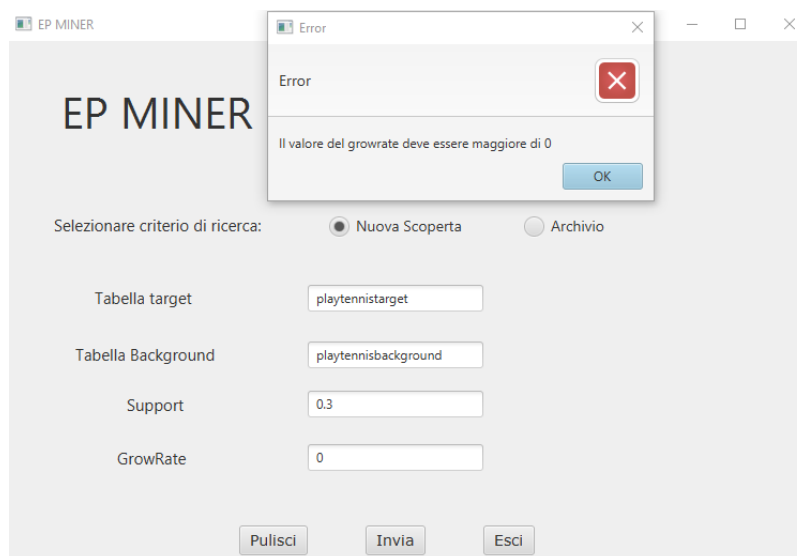
Input: dati tali da non avere emerging pattern

Output: stampa dei frequent pattern con un avviso che gli emerging pattern non sono stati individuati



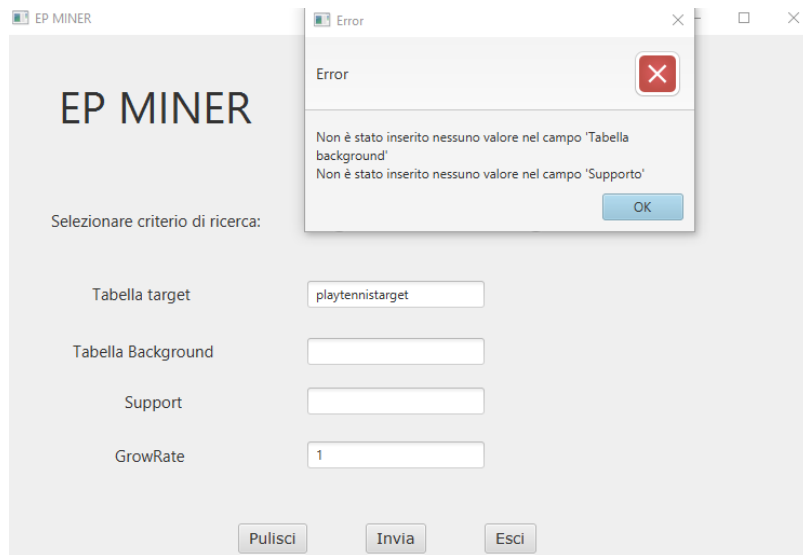
Input: valore di growrate minore o uguale a 0

Output: messaggio di errore relativo al growrate errato



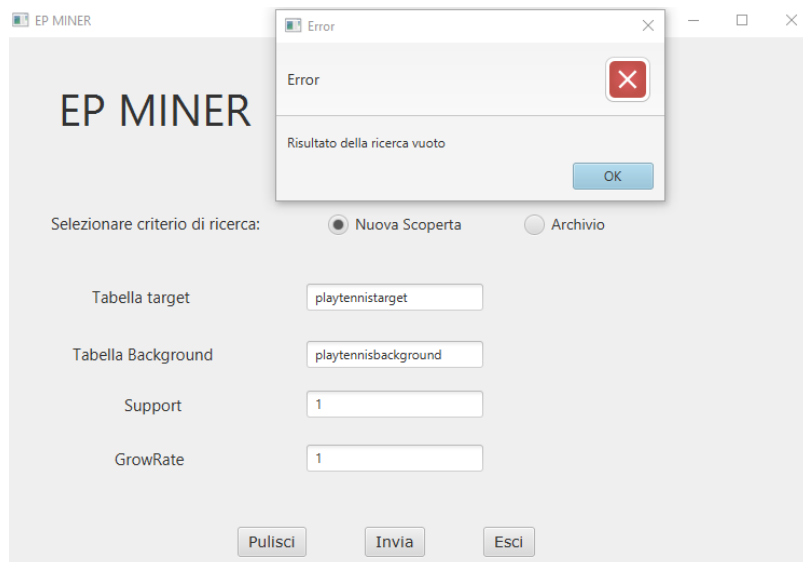
Input: campi vuoti dei dati relativi alla ricerca dei pattern

Output: messaggio di errore relativo ai dati non inseriti



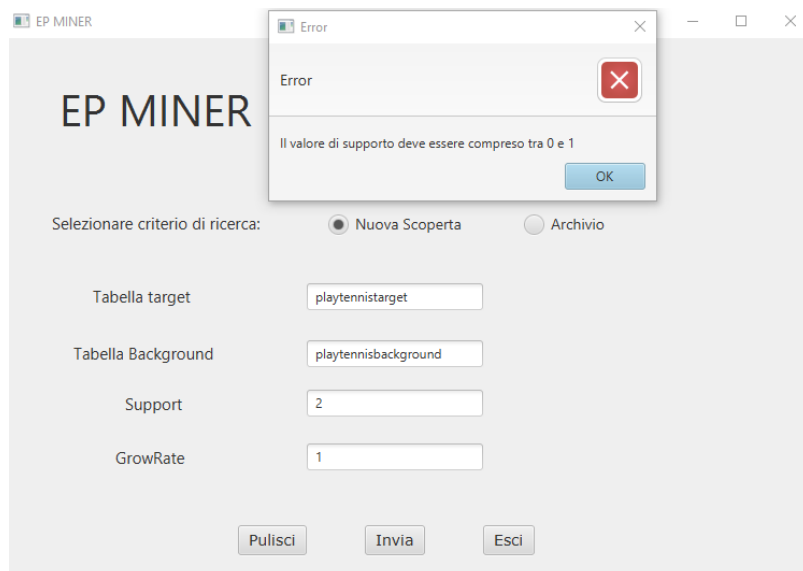
Input: valori tali da avere i frequent pattern e gli emerging pattern vuoti

Output: messaggio di errore relativo all'assenza di risultati nella ricerca



Input: valore di supporto maggiore di 1 o minore di 0

Output: messaggio di errore relativo al supporto errato



Input: valore non numerico nel campo growrate o nel campo del minimo supporto

Output: messaggio di errore relativo all'inserimento del valore non numerico

