

EpMiner: Scoperta di Pattern Frequenti e Pattern Emergenti

Caso di studio di Metodi Avanzati di Programmazione
AA 2020-2021



Realizzato da

Capone Simone 717216
s.capone7@studenti.uniba.it

Giacovazzo Christian 716696
c.giacovazzo1@studenti.uniba.it

Giannini Mariangela 680636
m.giannini16@studenti.uniba.it

INDICE

1. <u>Introduzione</u>	2
1.1 <u>Algoritmo Apriori</u>	2
1.2 <u>Limiti</u>	3
2. <u>Introduzione al progetto</u>	4
2.1 <u>Descrizione del progetto</u>	4
3. <u>Diagrammi UML</u>	5
3.1 <u>UML Packages</u>	5
3.2 <u>UML Classi</u>	8
4. <u>Guida all'installazione</u>	9
4.1 <u>Installazione Server</u>	9
4.2 <u>Installazione Client</u>	9
5. <u>Guida Utente</u>	10
5.1 <u>Guida alla interazione da console</u>	10
6. <u>Note</u>	14

1. Introduzione

1.1 Algoritmo Apriori

L'algoritmo Apriori è stato proposto da R. Agrawal e R. Srikant nel 1994.

Esso è progettato per operare su database contenenti transazioni. Ogni transazione è vista come un insieme di elementi. Data una soglia C , l'algoritmo Apriori identifica gli insiemi di elementi che sono sottoinsiemi di almeno C transazioni nel database.

Apriori utilizza un approccio "dal basso verso l'alto", in cui i sottoinsiemi frequenti vengono estesi un elemento alla volta e i gruppi di candidati vengono testati rispetto ai dati.

L'algoritmo termina quando non vengono trovate ulteriori estensioni riuscite.

Apriori utilizza la ricerca in ampiezza (*breadth-first search*) e una struttura ad albero Hash per contare in modo efficiente i set di elementi candidati. Genera insiemi di elementi candidati di lunghezza k da insiemi di elementi di lunghezza $k-1$. Quindi sfortisce i candidati che hanno un sottomodulo poco frequente. L'insieme candidato contiene così tutti gli insiemi di elementi frequenti di lunghezza k . Successivamente, esegue la scansione del database delle transazioni per determinare gli insiemi di elementi frequenti tra i candidati.

Lo pseudo codice per l'algoritmo è riportato di seguito. Viene impiegata la notazione della teoria degli insiemi. Ad ogni passo, si assume che l'algoritmo generi gli insiemi candidati dagli insiemi di elementi del livello precedente.

```
frequentPatternDiscovery(DTarget,minS)→FP
begin
    FP = ∅
    L_1 = {1 item che compaiono in minS x |D| transazioni di DTarget}
    K=2
    while L_{k-1} ≠ ∅ do
        begin
            C_k = {candidati generati da L_{k-1} aggiungendo un nuovo item}
            L_k = ∅
            for each (p e C_k) do
                if (supporto(p, DTarget) ≥ minS) then
                    L_k = L_k ∪ p
            FP = FP ∪ L_k
            K = K + 1
        end
    return FP
end

EPDiscovery(DBackground,FP,minGr)→EP
begin
    EP = ∅
    for each (p e FP) do
        begin
            if (growrate(p,DBacKground) ≥ minGR) then
                EP = EP ∪ p
        end
    return EP
end
```

1.2 Limiti

L'algoritmo Apriori soffre di una serie di inefficienze. La generazione di candidati genera un numero elevato di sottoinsiemi, infatti l'algoritmo tenta di caricare il gruppo di candidati, con il maggior numero possibile di sottoinsiemi prima di ogni scansione del database.

L'esplorazione dal basso verso l'alto del sottoinsieme trova qualsiasi sottoinsieme massimale S solo dopo tutti i $2^{|S|} - 1$ suoi sottoinsiemi propri.

L'algoritmo esegue la scansione del database troppe volte, il che riduce le prestazioni complessive. Per questo motivo, l'algoritmo presuppone che il database sia permanentemente in memoria.

Inoltre, sia la complessità temporale che spaziale di questo algoritmo sono molto elevate: $O(2^{|D|})$, quindi esponenziale, dove $|D|$ è il numero totale di elementi presenti nel database.

2. Introduzione al progetto

2.1 Descrizione del progetto

Il software realizzato utilizza l'algoritmo Apriori, descritto nella sezione precedente, elaborando dati estratti da una tabella presente in un database di tipo MySQL.

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

Il server si occupa di ricevere le richieste di uno o più client, i quali posso effettuare le seguenti operazioni:

- Ricerca di un nuovo pattern nel database
- Caricamento di un pattern salvato in archivio nel database.

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

- Minimo supporto
- Minimo grow rate
- Nome delle tabelle target e background

Il server mostra inoltre informazioni sul client connesso: le operazioni da esso richieste e il loro esito.

Nella sezione 3 sono riportati anche i diagrammi UML per i package e per le classi.

Inoltre nella cartella "*Javadoc*" è stata allegata la Javadoc creata direttamente dall'IDE di sviluppo (IntelliJ).

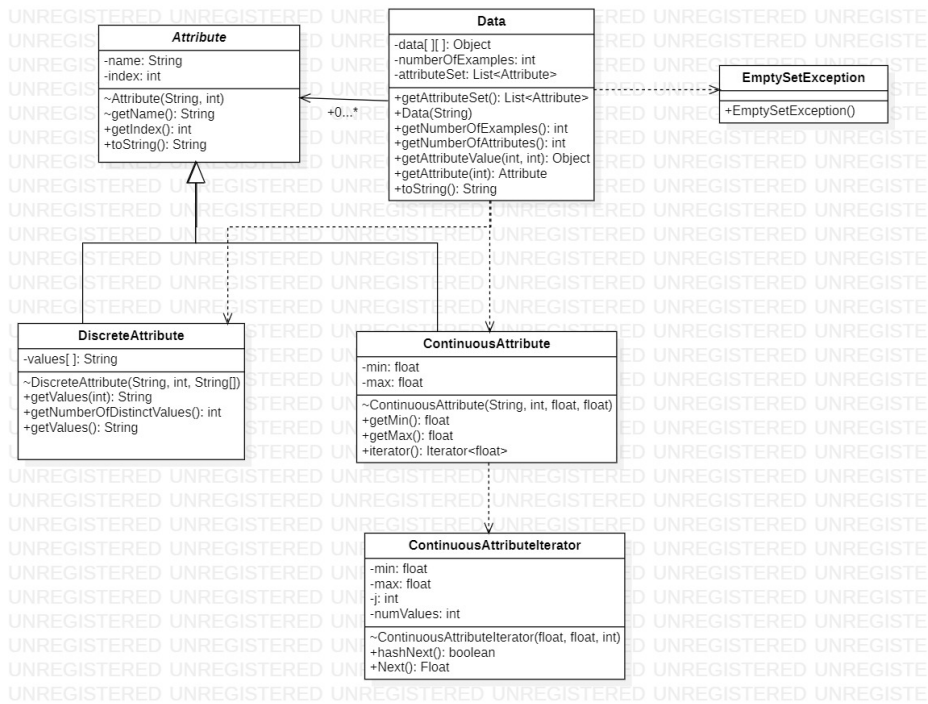
Nella sezione 5 del documento sono riportati esempi di esecuzione.

3. Diagrammi UML

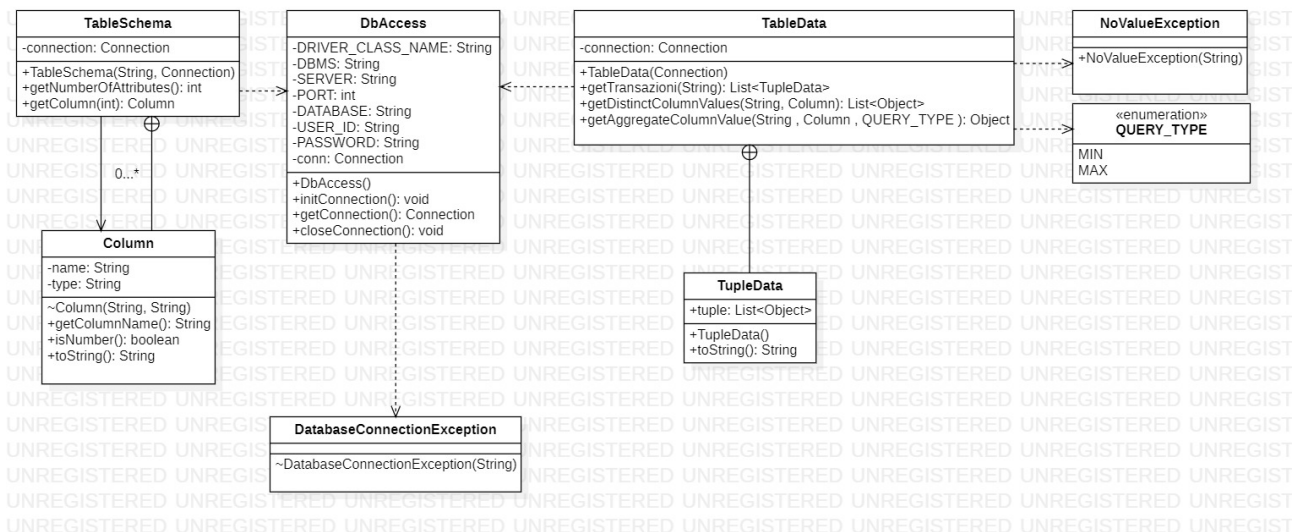
Per la realizzazione dei diagrammi è stato utilizzato il software *StarUML*.

3.1 UML Packages

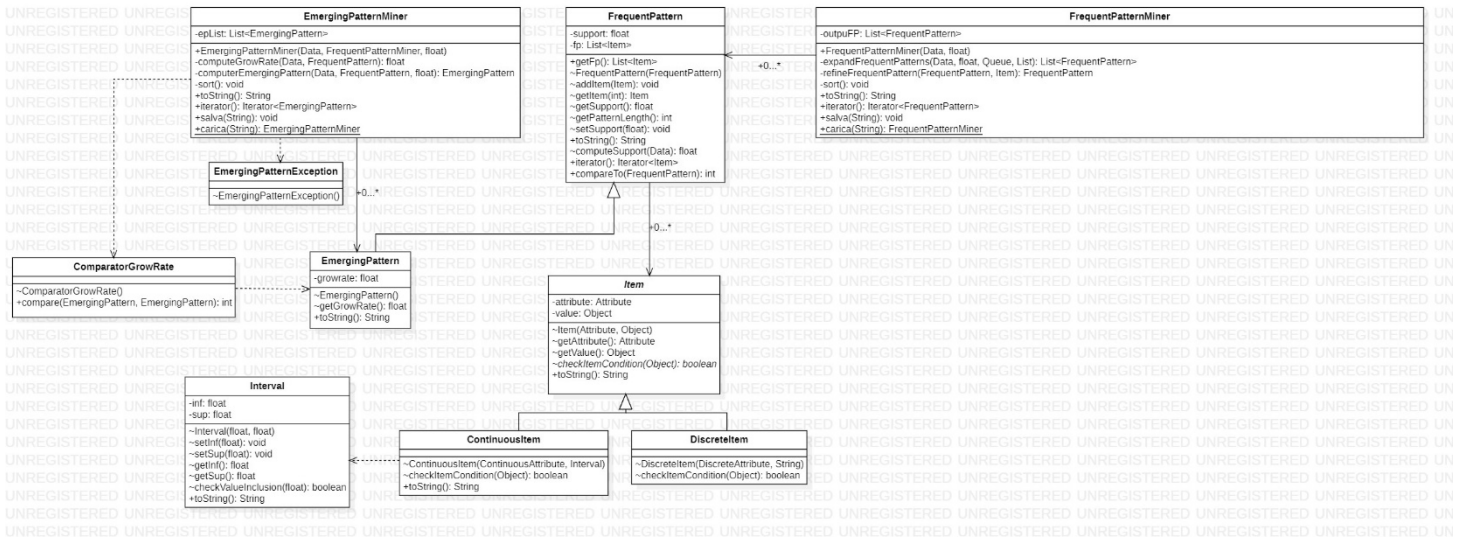
Package data



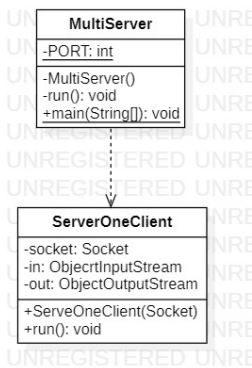
Package database



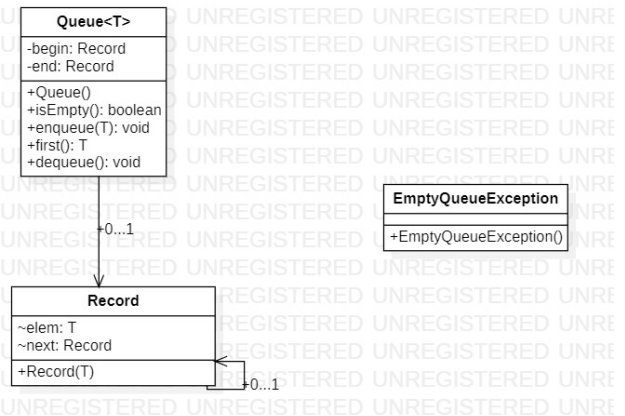
Package mining



Package server



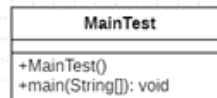
Package utility



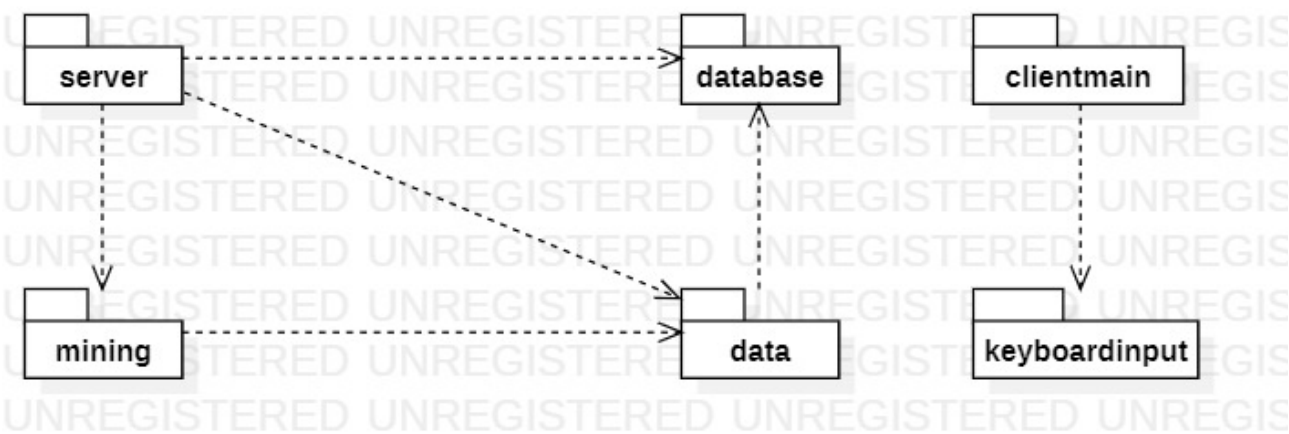
Package keyboardinput



Package clientmain



Interazione Packages



3.2 UML Classi

Per garantire una maggiore leggibilità, nel diagramma delle classi EpMinerServer_base sono stati omessi metodi e attributi.

Diagramma completo delle classi EpMinerServer_base

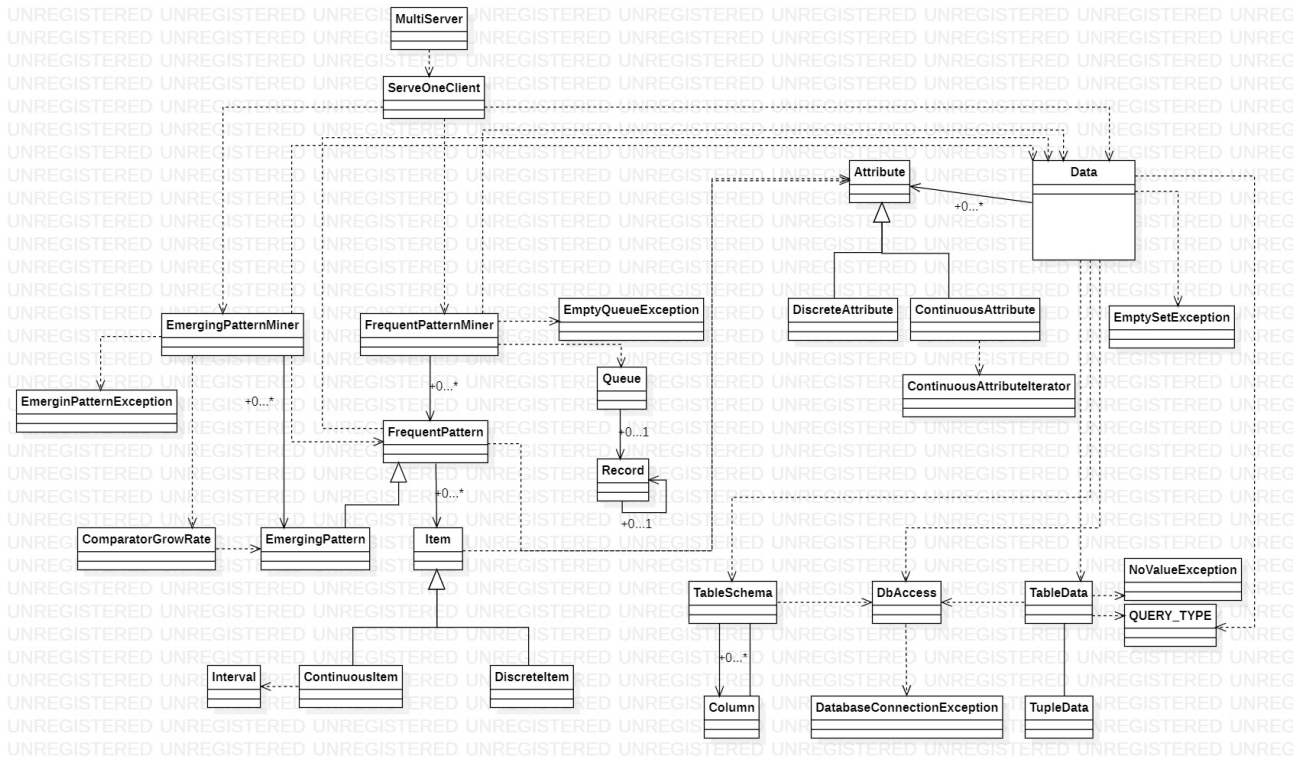
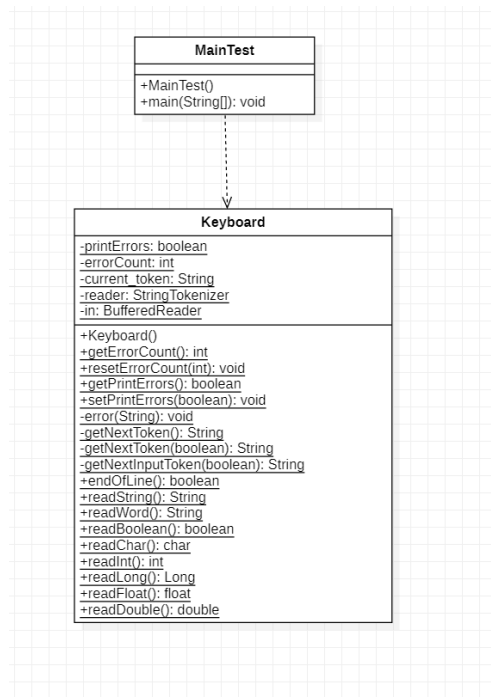


Diagramma delle classi EpMinerClient_base



4. Guida all'installazione

4.1 Installazione Server

Per il corretto funzionamento del progetto lato server è necessario:

- installare MySQL 8.0;
- installare Java Runtime Environment (JRE) versione 16;
- linkare al progetto relativo al server il file *"mysql-connector-java-8.0.17"*, responsabile della connessione al database MySQL.¹
Tale connettore è reperibile nella cartella *"SQL Connector e script"*;
- avviare il server MySQL;
- eseguire lo script MySQL presente nella cartella *"SQL Connector e script"*.
Tale script inizializza il database con tabelle e tuple di esempio.²

Per un più immediato utilizzo, la porta 3306 è preimpostata per accedere al database MySQL.

4.1 Installazione Client

Per il corretto funzionamento del progetto lato client è necessario:

- installare Java Runtime Environment (JRE) versione 16;
- avviare il server. In caso contrario ci sarà un messaggio di errore.

Per un più immediato utilizzo, l'indirizzo IP al quale il client si connette è preimpostato a quello locale 127.0.0.1 con porta 8080.

5. Guida Utente

Nella cartella principale è presente una sotto cartella “*EpMiner_Archive*”, nella quale verranno salvati (e caricati) in file *.bin* tutti i pattern trovati. Spostare tale cartella sul Desktop. In essa sono presenti già dei file a scopo di esempio.

Le tabelle di esempio presenti nello script *mySQL* si chiamano “*playtennistarget*” e “*playtennisbackground*”.

5.1 Guida base alla interazione da console

Nella cartella “*Eseguibile*” eseguire il file “*Server.bat*” e il file “*EpMinerClient_base.bat*”. Si apriranno due distinte schermate a linea di comando: una per il server e una per il client.

1) Avvio server:



```
Started: ServerSocket[addr=0.0.0.0/0.0.0.0,localport=8080]
```

2) Avvio Client:



```
Started: ServerSocket[addr=0.0.0.0/0.0.0.0,localport=8080]

Nuovo client connesso.
```



```
addr = /127.0.0.1
port=8080
Socket[addr=/127.0.0.1,port=8080,localport=56850]
Scegli una opzione:
1:Nuova scoperta
2: Risultati in archivio
```

A questo punto bisogna digitare da tastiera le informazioni richieste per la computazione.

3) Nuova scoperta:



```
Started: ServerSocket[addr=0.0.0.0/0.0.0.0,localport=8080]

Nuovo client connesso.
Elaborazione nuova richiesta...
Richiesta completata con successo.
```



```
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
Frequent patterns
1:outlook=rain[0.375]
2:outlook=sunny[0.375]
3:temperature in [0.0,6.06[[0.375]
4:temperature in [24.24,30.300098[[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 [0.0,6.06[ AND umidity=normal[0.375]
10:temperature in [24.24,30.300098[ AND umidity=high[0.375]
11:umidity=high AND outlook=sunny[0.375]
12:umidity=high AND temperature in [24.24,30.300098[[0.375]
13:umidity=high AND play=no[0.375]
14:umidity=normal AND temperature in [0.0,6.06[[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]
Emerging patterns
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 [0.0,6.06[[0.375][2.25]
6:temperature in [24.24,30.300098[[0.375][2.25]
7:temperature in [0.0,6.06[ AND umidity=normal[0.375][2.25]
8:umidity=high AND play=no[0.375][2.25]
9:umidity=normal AND temperature in [0.0,6.06[[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 [24.24,30.300098[ AND umidity=high[0.375][Infinity]
15:umidity=high AND outlook=sunny[0.375][Infinity]
16:umidity=high AND temperature in [24.24,30.300098[[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]
Vuoi ripetere?(s/n)
█
```

4) Risultati in archivio:



```
Vuoi ripetere?(s/n)
s
Scegli una opzione:
1:Nuova scoperta
2: Risultati in archivio
2
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
Frequent patterns
1:outlook=rain[0.375]
2:outlook=sunny[0.375]
3:temperature in [0.0,6.06[[0.375]
4:temperature in [24.24,30.300098[[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 [0.0,6.06[ AND umidity=normal[0.375]
10:temperature in [24.24,30.300098[ AND umidity=high[0.375]
11:umidity=high AND outlook=sunny[0.375]
12:umidity=high AND temperature in [24.24,30.300098[[0.375]
13:umidity=high AND play=no[0.375]
14:umidity=normal AND temperature in [0.0,6.06[[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]
Emerging patterns
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 [0.0,6.06[[0.375][2.25]
6:temperature in [24.24,30.300098[[0.375][2.25]
7:temperature in [0.0,6.06[ AND umidity=normal[0.375][2.25]
8:umidity=high AND play=no[0.375][2.25]
9:umidity=normal AND temperature in [0.0,6.06[[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 [24.24,30.300098[ AND umidity=high[0.375][Infinity]
15:umidity=high AND outlook=sunny[0.375][Infinity]
16:umidity=high AND temperature in [24.24,30.300098[[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]
Vuoi ripetere?(s/n)
█
```

In caso di assenza del file o tabella, verrà stampato un messaggio di avviso e sarà possibile effettuare una nuova operazione.

5) Casi particolari:



```
Started: ServerSocket[addr=0.0.0.0/0.0.0.0,localport=8080]
Nuovo client connesso.
Elaborazione nuova richiesta...
Tabella non trovata.
Elaborazione nuova richiesta...
File non trovato.
```



```
addr = /127.0.0.1
port=8080
Socket[addr=/127.0.0.1,port=8080,localport=54074]
Scegli una opzione:
1: Nuova scoperta
2: Risultati in archivio
1
Inserire valore minimo supporto (minsup>0 e minsup<=1):
0.7
Inserire valore minimo grow rate (minGr>0):
5
Tabella target:
pippo
Tabella background:
topolino
Impossibile elaborare la richiesta.

Non è stata trovata una tabella avente il nome da lei indicato.
Vuoi ripetere?(s/n)
s
Scegli una opzione:
1: Nuova scoperta
2: Risultati in archivio
2
Inserire valore minimo supporto (minsup>0 e minsup<=1):
0.3
Inserire valore minimo grow rate (minGr>0):
1
Tabella target:
p
Tabella background:
p
Impossibile elaborare la richiesta.

Non è stata trovata nessuna tabella in archivio con i dati da lei inseriti.
Vuoi ripetere?(s/n)
```

Per uscire dal programma, basterà digitare “n” quando richiesto oppure chiudere direttamente la schermata. Per spegnere il server chiudere la relativa schermata.

6. Note

¹ *Nel caso in cui il file jar non funzionasse a causa del connettore non linkato correttamente, è possibile importare quest'ultimo come libreria esterna al progetto, ed avviare il server direttamente da IDE.*

² *In alternativa si può aprire il file con un editor di testo e copiare il contenuto nella Shell MySQL.*