

Interrogazioni con preferenze su domini tassonomici

Candidato: Dragos Mihaita Iftimie

Relatore: Prof. Paolo Ciaccia

Società dell'informazione

Il trattamento e l'elaborazione dell'informazione è diventata la principale fonte occupazionale. Informazione e conoscenza sono le materie prime e i principali prodotti della nostra società.

Passaggio da un numero di informazioni notevoli, ma gestibili, a una quantità enorme che ci raggiunge con qualsiasi mezzo.

Blocco decisionale

Query con preferenze

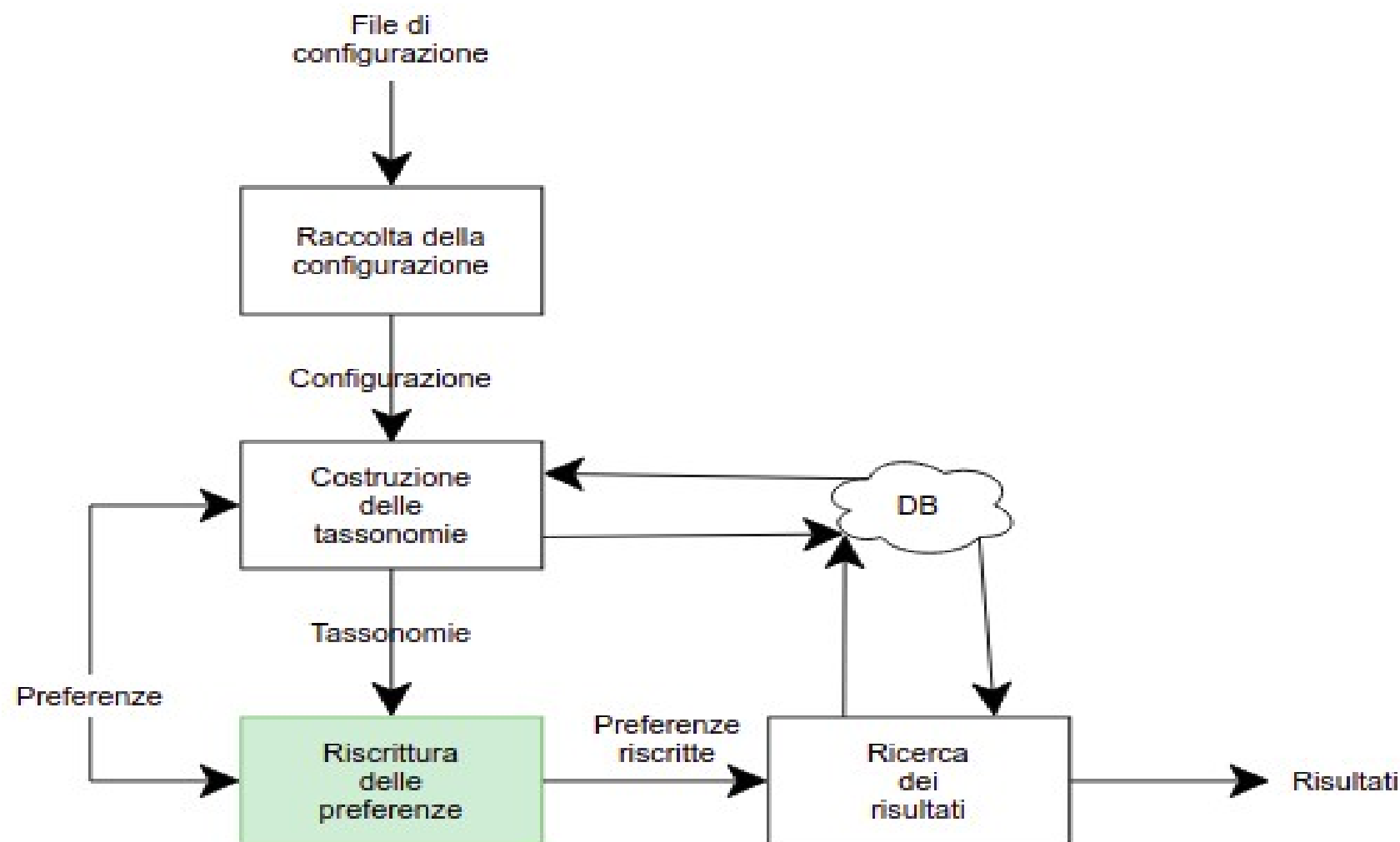
Limitazioni:

- Preferenze a un livello più generico rispetto ai dati tra cui si cerca;
- Contraddizioni tra preferenze dovute al loro differente livello di dettaglio;
- Mancanza di transitività tra preferenze.

Soluzioni:

- Domini tassonomici definiti sugli attributi dei dati presenti;
- Riscrittura delle preferenze tramite l'operatore S;
- Riscrittura delle preferenze tramite l'operatore T.

Scopo della tesi



Appunti necessari

Tassonomia QueryBased: tassonomia definita in un database;

Tassonomia RuleBased: tassonomia definita a partire da una regola;

Parte dinamica di una tassonomia: insieme delle foglie della tassonomia, ogni foglia è un valore appartenente ai dati che l'utente sta cercando;

Parte statica di una tassonomia: insieme dei nodi di una tassonomia non dinamici.

Raccolta della configurazione

- informazioni per la comunicazione con il database

```
{  
  "connection":{  
    "host":"192.168.56.1",  
    "database":"SimplePreference",  
    "user":"sample",  
    "password":"samplepass"  
  }  
}
```

- informazioni per la costruzione di tassonomie memorizzate nel database

```
mysql> select * from artistTax;  
+-----+-----+  
| moreSpecific | moreGeneric |  
+-----+-----+  
| Bruce Springsteen | rock |  
| Paul McCartney | rock |  
| Eminem | rap |  
| Trap God | trap |  
| Rihanna | pop |  
| Rihanna | rap |  
| Madonna | pop |  
| Machine Gun Kelly | rap |  
| Machine Gun Kelly | trap |  
| Machine Gun Kelly | punk |  
+-----+-----+
```

```
{  
  "attribute":"artist",  
  "fixed":  
    [  
      {  
        "template":"flatTemplate",  
        "attribute":"moreGeneric",  
        "table":"artistTax"  
      }  
    ],  
  "dynamic":  
    [  
      {  
        "template":"sameTable",  
        "moreSpecific":"moreSpecific",  
        "moreGeneric":"moreGeneric",  
        "table":"artistTax"  
      }  
    ]  
}
```

Raccolta della configurazione

- informazioni per la costruzione di tassonomie definite attraverso regole

```
"intervals":  
  [  
    {  
      "value": "winter",  
      "lowerInterval":  
        [  
          "-inf", "12", "23"  
        ],  
      "upperInterval":  
        [  
          "+inf", "3", "20"  
        ],  
      "lowerInclusion":  
        [  
          "["  
        ],  
      "upperInclusion":  
        [  
          "]"  
        ]  
    },  
  ],
```

```
    "value": "autumn",  
    "lowerInterval":  
      [  
        "-inf", "9", "23"  
      ],  
    "upperInterval":  
      [  
        "+inf", "12", "22"  
      ],  
    "lowerInclusion":  
      [  
        "["  
      ],  
    "upperInclusion":  
      [  
        "]"  
      ]  
  },
```

```
    "value": "spring",  
    "lowerInterval":  
      [  
        "-inf", "3", "21"  
      ],  
    "upperInterval":  
      [  
        "+inf", "6", "21"  
      ],  
    "lowerInclusion":  
      [  
        "["  
      ],  
    "upperInclusion":  
      [  
        "]"  
      ]  
  },  
  {  
    "value": "summer",  
    "lowerInterval":  
      [  
        "-inf", "6", "22"  
      ],  
    "upperInterval":  
      [  
        "+inf", "9", "22"  
      ],  
    "lowerInclusion":  
      [  
        "["  
      ],  
    "upperInclusion":  
      [  
        "]"  
      ]  
  }  
]
```

Costruzione delle tassonomie

Considerando le necessità degli operatori T ed S, si costruisce una tassonomia nel seguente modo:

- costruzione della parte statica
- indicazione minima delle intersezioni dinamiche tra nodi statici
- inserimento dei nodi dinamici presenti nelle preferenze

```
F2(x,y): x[artist] ≤ Machine Gun Kelly & y[artist] ≤ pop
```

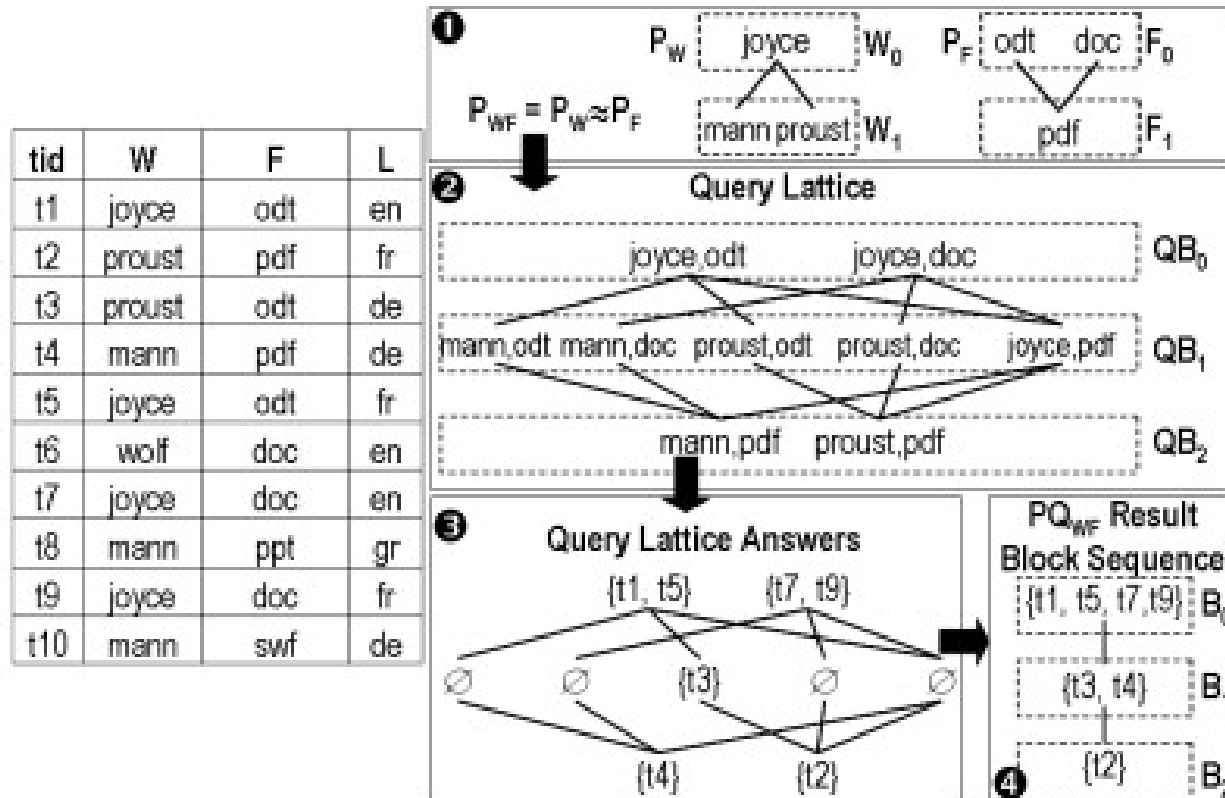
```
LEVEL 1:topValue  
LEVEL 2:rock -- punk -- trap -- rap -- pop  
LEVEL 3:PLACEHOLDER1 -- PLACEHOLDER0
```

```
92  
PLACEHOLDER0:pop.  
PLACEHOLDER0:rap.  
PLACEHOLDER1:punk.  
PLACEHOLDER1:rap.  
PLACEHOLDER1:trap.
```

```
LEVEL 1:topValue  
LEVEL 2:pop -- punk -- trap -- rap -- rock  
LEVEL 3:PLACEHOLDER0 -- PLACEHOLDER1 --  
Machine Gun Kelly
```

```
163  
Machine Gun Kelly:punk.  
Machine Gun Kelly:rap.  
Machine Gun Kelly:trap.  
PLACEHOLDER0:pop.  
PLACEHOLDER0:rap.  
PLACEHOLDER1:punk.  
PLACEHOLDER1:rap.  
PLACEHOLDER1:trap.
```


Ricerca dei risultati



LBA adattato ricordando:

- non pensato per funzionare su domini tassonomici;
- possibili preferenze del tipo:
 $x[\text{writer}] \leq \text{joyce} \ \& \ x[\text{language}] \leq \text{english} \ \& \ y[\text{writer}] \leq \text{mann} \ \& \ y[\text{language}] \leq \text{italian};$
- possibili preferenze del tipo:
 $x[\text{writer}] \leq \text{joyce} \ \& \ y[\text{format}] \leq \text{odt}.$

Ricerca dei risultati - preferenze

```
UserPrefs
F1(x,y): x[price] ≤ cheap & y[price] ≤ expensive
F2(x,y): x[artist] ≤ Machine Gun Kelly & y[artist] ≤ pop
F3(x,y): x[venue] ≤ InDoor & x[day] ≤ winter & y[venue] ≤ OutDoor & y[day]
          ≤ winter
```

```
PREFERENCES UserPrefsSTST:
p1
  F1(x,y): x[price] ≤ cheap & y[price] ≤
           expensive
p2
  F2(x,y): x[artist] ≤ Machine Gun Kelly &
           y[artist] ≤ pop
p3
  F3(x,y): x[day] ≤ winter & x[venue] ≤ InDoor
           & y[day] ≤ winter & y[venue] ≤ OutDoor
p4 //T:p1+p2
  c49(x,y): x[price] ≤ cheap & y[artist] ≤ pop
           //T:F1+F2
p5 //T:p1+p3
  c50(x,y): x[price] ≤ cheap & y[day] ≤ winter
           & y[venue] ≤ OutDoor //T:F1+F3
p6 //T:p2+p1
  c51(x,y): x[artist] ≤ Machine Gun Kelly &
           y[price] ≤ expensive //T:F2+F1
p7 //T:p2+p3
  c52(x,y): x[artist] ≤ Machine Gun Kelly &
           y[day] ≤ winter & y[venue] ≤ OutDoor
           //T:F2+F3
p8 //T:p3+p1
  c53(x,y): x[day] ≤ winter & x[venue] ≤ InDoor
           & y[price] ≤ expensive //T:F3+F1
p9 //T:p3+p2
  c54(x,y): x[day] ≤ winter & x[venue] ≤ InDoor
           & y[artist] ≤ pop //T:F3+F2
```

Ricerca dei risultati - migliori risultati

```
select distinct price,day,id,venue,artist
from concerts where artist in
('Madonna','Rihanna') AND price in
('30','35','25','50') AND day in
('2019-02-01','2020-01-05') AND venue in
('Blue Note');
select distinct price,day,id,venue,artist
from concerts where venue in ('Blue
Note') AND artist in ('Machine Gun
Kelly') AND price in
('30','35','25','50') AND day in
('2019-02-01','2020-01-05');
select distinct price,day,id,venue,artist
from concerts where artist in
('Madonna','Rihanna') AND day in
('2019-02-01','2020-01-05') AND price in
('1000','100','120','140') AND venue in
('Unipol arena','Stadio olimpico','Verona
Arena');
select distinct price,day,id,venue,artist
from concerts where venue in ('Unipol
arena','Stadio olimpico','Verona Arena')
AND price in ('30','35','25','50') AND
artist in ('Madonna','Rihanna') AND day
in ('2019-02-01','2020-01-05');
select distinct price,day,id,venue,artist
from concerts where day in
('2019-02-01','2020-01-05') AND venue in
('Blue Note') AND artist in ('Machine Gun
Kelly') AND price in
('1000','100','120','140');
select distinct price,day,id,venue,artist
from concerts where price in
('1000','100','120','140') AND artist in
('Machine Gun Kelly') AND day in
('2019-02-01','2020-01-05') AND venue in
('Unipol arena','Stadio olimpico','Verona
Arena');
```

```
mysql> select * from concerts;
```

id	artist	day	venue	price
1	Bruce Springsteen	2019-05-10	Verona Arena	70
2	Madonna	2019-06-24	Verona Arena	35
3	Madonna	2019-07-21	Blue Note	120
4	Eminem	2019-08-12	Unipol arena	60
5	Rihanna	2019-10-10	Blue Note	50
6	Bruce Springsteen	2019-10-30	Stadio olimpico	100
7	Paul McCartney	2019-11-01	Stadio olimpico	1000
8	Trap God	2019-05-11	Unipol arena	30
9	Machine Gun Kelly	2020-01-05	Unipol arena	25
10	Rihanna	2019-02-01	Blue Note	80

```
select distinct price,day,id,venue,artist
from concerts where venue in ('Unipol
arena','Stadio olimpico','Verona Arena')
AND price in ('30','35','25','50') AND
artist in ('Machine Gun Kelly') AND day
in ('2019-02-01','2020-01-05');
select distinct price,day,id,venue,artist
from concerts where artist in
('Madonna','Rihanna') AND day in
('2019-02-01','2020-01-05') AND venue in
('Blue Note') AND price in
('1000','100','120','140');
level 0: [[25, 2020-01-05, 9, Unipol arena,
Machine Gun Kelly]]
level 1: [[140, 2019-02-01, 11, Blue Note,
Rihanna]]
```

Fine