TD 4: Stratification and Query Optimization

Exercise 1. (Query Optimization)

Let us consider the following database:
EDB (schema)

```
emp(Ssn,Salary,Num_Dep,Age).
dept(Num_Dep,Ssn_Resp,Floor).
sales(Num_Dep,Article,Vol).
```

The IDB and the constraints are as follows:
IDB (Rules):

```
highDepSales(x1,x2,x3,y2,y3) :- dept(x1,x2,x3),sales(x1,y2,y3),y3>100000.
highMgrSales(x2,y2,y4) :- emp(x2,y2,y3,y4),highDepSales(x1,x2,x3,x4,x5).
```

IC:

```
[IC1] ⊥:- dept(x,y,2)
[IC2] (y>40000):- emp(x,y,z,u), (u>50).
```

Questions: Answer the following questions:

1. What do the constraints IC3 and IC4 mean?
   Solution:
   IC1: There is no department on floor 2.
   IC2: All employees over the age of 50 have a salary above 40,000.

2. Rewrite the IDB clauses using only basic relations in the body of the clauses.
   Solution:

```
highMgrSales(x2,y2,y4) :- emp(x2,y2,y3,y4),dept(x1,x2,x3),
                          sales(x1,z2,z3), z3>100000.
```

3. Consider the following query:

```
Answer(x,z) :- highDepSales(x,y,2,z,u).
```

   What does the evaluation of this request return by considering the EDB, IDB and IC?
   Solution:
   An empty set.

4. Consider the following query:

```
Answer(x1,x2) :- highMgrSales(x1,x2,x3)(x3>50).
```

   Rewrite this query by considering the EDB, IDB and IC?
   Solution:

```
Answer(x2,y2) :- emp(x2,y2,y3,y4),dept(x1,x2,x3),
                 sales(x1,z2,z3), z3>100000, y4>50.
```

5. Provide the result of the compilation of EDB and IDB as well as the expansion of ICs.

Solution:

A1: `emp(x,y,z,u):- emp(x,y,z,u).`

A2: `dept(x,y,z):- dept(x,y,z).`

A3: `sales(x,y,z,u):- sales(x,y,z,u).`

A4:
```
highDepSales(x1,x2,x3,y2,y3) :-
                    dept(x1,x2,x3),sales(x1,y2,y3),y3>100000.
```

A5:
```
highMgrSales(x2,y2,y4) :- emp(x2,y2,y3,y4),dept(x1,x2,x3),
                    sales(x1,z2,z3), z3>100000.
```

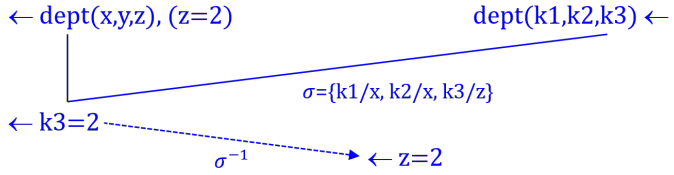
IC1+: `⊥ :- dept(x,y,z),(z=2).`

IC2+: `(y>40000):- emp(x,y,z,u),(u>50).`

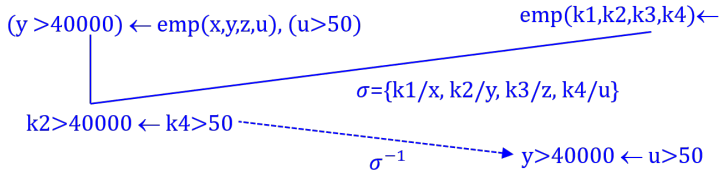
6. Provide the semantically constrained axioms (SCA).

Solution:

Refusion Tree for IC1+



Refusion Tree for IC2+



SCA1:
```
emp(x,y,z,u) :- emp(x,y,z,u){(y>40000)←(u>50)}
      emp(x,y,z,u) :- emp(x,y,z,u), not((u>50), not(y>40000))
```

SCA2:
```
dept(x,y,z) :- dept(x,y,z){⊥ ← z=2}
      dept(x,y,z) :- dept(x,y,z), not(z=2)
```

SCA3: `sales(x,y,z,u) :- sales(x,y,z,u){}`

SCA4:
```
highDepSales(x1,x2,x3,y2,y3) :-
          dept(x1,x2,x3),sales(x1,y2,y3),y3>100000{⊥ ← x3=2}.

      highDepSales(x1,x2,x3,y2,y3) :-
          dept(x1,x2,x3),sales(x1,y2,y3),y3>100000, not(x3=2).
```

SCA5:
```
highMgrSales(x2,y2,y4) :- emp(x2,y2,y3,y4),dept(x1,x2,x3),
          sales(x1,z2,z3), z3>100000{(y2>40000)←(y4>50), ⊥←x3=2}.

    highMgrSales(x2,y2,y4) :- emp(x2,y2,y3,y4),dept(x1,x2,x3),
    sales(x1,z2,z3), z3>100000, not((y4>50), not(y2>40000)), not(x3=2).
```

Exercise 2. (Datalog with Negation)

Indicate the correct rules and incorrect rules among the following:

```
1.  S(x) :- ¬R(x).
2.  S(x) :- R(x), x > y.
3.  S(x) :- S(x).
4.  S(x) :- R(x), ¬S(x).
```

Solution:

1. incorrect
2. incorrect
3. correct
4. correct, but it is unstratifiable.

Exercise 3. (Stratification)

Let us consider the following Datalog program Π:

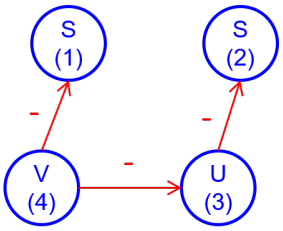
```
1. S(x) :- R1(x), ¬R(x).
2. T(x) :- R2(x), ¬R(x).
3. U(x) :- R3(x), ¬T(x):
4. V(x) :- R4(x), ¬S(x), ¬U(x).
```

Among the following propositions, indicate which are correct stratifications:

```
A. {1},{2},{3},{4}
B. {2},{1},{3},{4}
C. {2},{3},{1},{4}
D. {1,2},{3},{4}
E. {2},{1,3},{4}
F. {3},{1,2},{4}
G. {2,3},{1},{4}
```

Solution:

A. yes
B. yes
C. yes
D. yes
E. yes
F. no
G. no



Exercise 4. (Stratification)

Let us consider the following Datalog program Π:

$$\begin{cases} p \leftarrow \neg q \\ q \leftarrow \neg p \\ r \leftarrow r \end{cases}$$

Can we stratify this program?

Solution: No!