Homework #5 (11.26)

1、已知有关系模式 R(A, B, C)和 S(B, C, D)，每个属性都占 10 个字节，请估计下面的逻辑查询计划的 T(U), S(U)以及结果关系中每个属性的 V 值（假设满足"Containment of Value Sets"，并且选择条件中的值都在关系中存在）：

$$U=\pi_{AD}[(\sigma_{A=3 \wedge B=5} R) \bowtie S)]$$

相应的统计量如下：

```
T(R) = 100000,    V(R,A) = 20,     V(R,B) = 50,     V(R,C) = 150
T(S) = 5000,      V(S,B) = 100,    V(S,C) = 200,    V(S,D) = 30
```

2、已知有下面的 3 个基本表：

```
Movies (title, year)              -- 电影表
Actors (actorID, name)            -- 演员表
Acted_in (actorID, title, year)     -- 演出表
```

3 个基本表相关的统计量如下：

T(Movies)=50,000　　T(Actors)=200,000　　T(Acted_in)=1,000,000

V(Movies, title)=30,000　　V(Movies, year)=90　　V(Actors, actorID)=200,000

V(Actors, name)=160,000　　V(Acted_in, actorID)=180,000

V(Acted_in, title)=29,000　　V(Acted_in, year)=90

请估计下面查询的结果关系大小（精确到小数点后 3 位）：

1）SELECT * FROM Movies WHERE year=2000 AND title='The Killer'

2）SELECT * FROM Movies,Acted_in WHERE Movies.title= Acted_in.title AND Movies.year=Acted_in.year

[考虑两种情况：（title, year）是 Movies 的主键，（title, year）不是 Movies 的主键]

解：

1、假设R和S中每个属性的值均匀分布

S(U) = 20 bytes

$$T(\sigma_{A=3 \wedge B=5}(R)) = \frac{T(R)}{V(R,A) \times V(R,B)} = \frac{100000}{20 \times 50} = 100$$

$$T(\sigma_{A=3 \wedge B=5}(R) \bowtie S) = \frac{T(\sigma_{A=3 \wedge B=5}(R)) \cdot T(S)}{V(S,B) \cdot V(S,C)} = \frac{100 \times 5000}{100 \times 200} = 25$$

∵投影操作不影响元组数. ∴ $T(U)=25$

∵属性A在R中做了选择操作且S中无属性A. ∴ $V(U,A)=1$

根据集值保持的假设, $V(U,D)=V(S,D)=30$

综上, $S(U)=20$ bytes, $T(U)=25$, $V(U,A)=1$, $V(U,D)=30$

2. 将SQL语句转换为关系代数:

1) $W_1=\sigma_{year=2000\ AND\ title='The\ killer'}(Movies)$

2) $W_2=Movies \bowtie_{Movies.title=Acted\_in.title\ AND\ Movies.year=Acted\_in.year} Acted\_in$

· 若(title, year)是Movies的主键:

$T(W_1)=1$ (也有可能查不到, 则$T(W_1)=0$)

∵(title, year)是Movies的主键, Acted_in的外键

∴ $T(W_2)=T(Acted\_in)=1000000$

· 若(title, year)不是Movies的主键:

$$T(W_1)=\frac{T(Movies)}{V(Movies,year)\cdot V(Movies,title)}=\frac{50000}{90\times 30000}=0.019$$

$$T(W_2)=\frac{T(Movies)\cdot T(Acted\_in)}{V(Movies,title)\cdot V(Movies,year)}=\frac{50000\times 1000000}{30000\times 90}=18518.519$$