Query Optimization

2. Exercise

Johannes Merkle
Ralf Vogler

7. Mai 2012

Exercise 1

Part 1

The expressions are not equal. Example:

R	
A	B
a_1	b_1
a_1	b_2
a_2	b_1

S	
A	B
a_1	b_1

results in

$\prod_A(R-S)$
A
a_1
a_2

$\prod_A(R) - \prod_A(S)$s
A
a_2

Part 2

R	
A_R	A_J
1	x
2	a

S	
A_S	A_J
3	a

T	
A_T	A_J
4	x
5	a

results in:

$(R \rtimes S) \rtimes T$			
A_R	A_J	A_S	A_T
1	x		4
2	a	3	5

$R \rtimes (S \rtimes T)$			
A_R	A_J	A_S	A_T
1	x		
2	a	3	5

Exercise 2

Part 1

If $R1.x$ is a key, we can have at most 1 entry, as keys are unique. In this case, the selectivity is $\frac{1}{|R1|}$. If it is not a key, a way to estimate the selectivity is to assume uniform distribution of values of the domain and therefore the selectivity can be estimated as $\frac{|R1.x|}{|R1|}$ where $|R1.x|$ denotes the number of values of the domain.

Part 2

Given an estimation for the selectivity of $\sigma_{R1.x=c}$, we can estimate the selectivity of $\bowtie_{R1.x=R2.y}$ as $selectivity(\sigma_{R1.x=c}) * selectivity(\sigma_{R1.x=c})$, though this would not be very accurate.

Exercise 3

see files in folder *tinydb*