Chapter 3.5 Selectivity of operations

Estimation of the size of intermediate results for execution plan:

$$sel(\sigma_P R) = \frac{|\sigma_P R|}{|R|} \quad \Rightarrow \quad |\sigma_P R| = sel * |R|$$

Special cases for key comparisons, constant comparisons, interval search.

$$sel(R \bowtie S) = \frac{|R \bowtie S|}{|R \times S|} = \frac{|R \bowtie S|}{|R| * |S|}$$

$$\Rightarrow |R \bowtie S| = sel. * |R| * |S|$$

$$sel(\pi_A R) = \frac{|\pi_A R|}{|R|} \quad \Rightarrow$$

$$sel(R \div S) = \frac{|R \div S|}{|R|} \leq \frac{1}{|S|}$$

$$sel(R \cup S) = \frac{|R \cup S|}{|R| + |S|}$$

$$sel(R \cap S) = \frac{|R \cap S|}{|R|}$$

Note: $sel(R \overline{\cup} S) = 1$
$sel(\overline{\pi} R) = 1$

Distinguish: selectivity $\neq$ costs
Deliberation: selectivity of n : 1 Joins? Selecitivites as extensions to systemtables, for foreign keys!

Estimation methods:

- Distributions of attributes
- histograms
- evaluate once and measure
- observe and remember
- sampling techniques

$\Rightarrow$ rather large uncertainties, see book of Kemper

Propagation of Selectivities
in complex expressions

Lemma 1: $sel(\sigma_Q \sigma_P) = sel(\sigma_Q) * sel(\sigma_P)$
for uniform distributions and independence of attr. values, e.g. $Q \equiv R.A = a$
$P \equiv R.B = b$

Lemma 2:

$sel(R \bowtie S \bowtie T) = sel(R \bowtie S) * sel(S \bowtie T)$

Proof: let $jp_{R,S} = \frac{|R \bowtie S|}{|R|} = sel(_{R \bowtie S}) * |S|$
= number of join-partners from S for R-tuple.

$$sel(R \bowtie S \bowtie T) = \frac{|R \bowtie S \bowtie T|}{|R| * |S| * |T|} = \frac{|R| * jp_{R,S} * jp_{S,T}}{|R| * |S| * |T|}$$

$$= \frac{|R| * sel(R \bowtie S) * |S| * sel(S \bowtie T) * |T|}{|R| * |S| * |T|}$$

$$= sel(R \bowtie S) * sel(S \bowtie T)$$

Problem: uniform distribution and independence of attr. values often not existing, e.g. Profs., Mitarb., Alter

Lemma 3: let P be selection predicate on R,
i.e. $\sigma_P(R \bowtie S) = (\sigma_P R) \bowtie S$
$\Rightarrow sel(\sigma_P(R \bowtie S)) = sel(\sigma_P R)$

Bew: $sel(\sigma_P(R \bowtie S)) = \frac{|\sigma_P(R \bowtie S)|}{|R \bowtie S|}$

$$= \frac{|\sigma_P(R \bowtie S)|}{|R| * |S| * sel(R \bowtie S)} = \frac{|(\sigma_P R) \bowtie S|}{|R| * |S| * sel(R \bowtie S)}$$

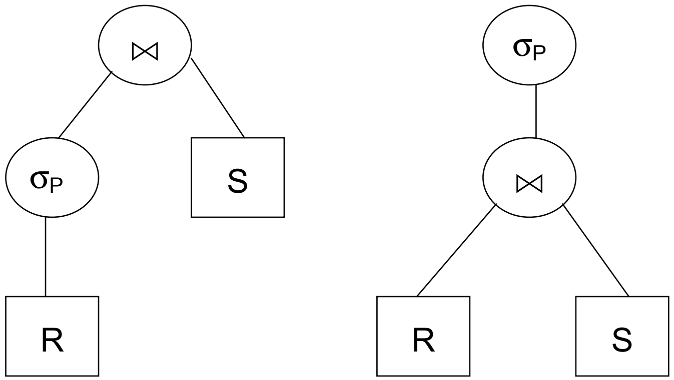
$$= \frac{sel(\sigma_P R) * |R| * jp_{R,S}}{|R| * |S| * sel(R \bowtie S)}$$

...

$$= \frac{sel(\sigma_P R) * |R| * sel(R \bowtie S) * |S|}{|R| * |S| * sel(R \bowtie S)} = sel(\sigma_P R)$$

Lemma 4:

sel $((\sigma_p \circ \bowtie)(R, S))$ = sel $(\sigma_P R)$ * sel $(R \bowtie S)$



Bew: Exercise

General Strategy:

1. Algebraic Trafos ⇒ many semantically equivalent operator trees

2. ordering properties and availability of indexes, key properties
   ⇒ choice of specific algorithms
   ⇒ specific operator-trees

3. selectivities ⇒ estimation of size of intermediate results, no costs yet!