

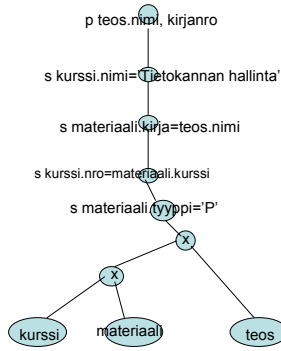


## Kyselypuut ja ekvivalenssi

Sisäisessä esityksessä kyselyt esitetään kyselypuuna

- lehdissä taulut
- juuressa lopputulos
- välisolmuina suoritettavat operaatiot

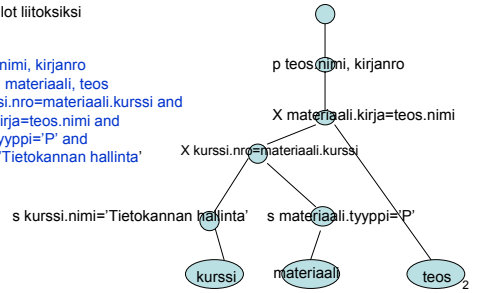
```
select teos.nimi, kirjanro
from kurssi, materiaali, teos
where kurssi.nro=materiaali.kurssi and
materiaali.kirja=teos.nimi and
materiaali.tyyppi='P' and
kurssinimi='Tietokannan hallinta'
```



## Kyselypuut ja ekvivalenssi

Optimoitu kyselypuu  
 • valinnat mahdollisimman aikaisin  
 • ristitulot liittokseksi

```
select teos.nimi, kirjanro
from kurssi, materiaali, teos
where kurssi.nro=materiaali.kurssi and
materiaali.kirja=teos.nimi and
materiaali.tyyppi='P' and
kurssinimi='Tietokannan hallinta'
```



## Kyselypuut ja ekvivalenssi

- Kyselypuun muokkauksessa vaihdetaan operaatioiden järjestyksiä tai korvataan jokin operaatio toisella operaatiolla
- Muokatun kyselypuun pitäisi tuottaa sama tulos kuin aiemman
- Kaksi operaatiosarjaa ovat **ekvivalentteja**, jos ne tuottavat saman tuloksen lähtötaulujen kaikilla instansseilla
  - ekvivalenssisäännöt määrittelevät millaiset muunnokset ovat sallittuja ilman, että tulos muuttuu

3



## Kyselypuut ja ekvivalenssi

- Valinnat voidaan yhdistää tai purkaa sisäkkäisiksi
 
$$\sigma_{c_1} \wedge \sigma_{c_2} \wedge \dots \wedge \sigma_{c_k} (R) \equiv \sigma_{c_1}(\sigma_{c_2}(\dots \sigma_{c_k}(R) \dots))$$
- Valinnat ovat vaihdannaisia (commutative)
 
$$\sigma_{c_1}(\sigma_{c_2}(R)) \equiv \sigma_{c_2}(\sigma_{c_1}(R))$$
- Peräkkäisten projektioiden lopputuloksen määrä viimeisin projektio
 
$$\pi_{a_1}(R) \equiv \pi_{a_1}(\pi_{a_2}(\dots(R)))$$

4



## Kyselypuut ja ekvivalenssi

- Ristitulon ja liittoksen ovat kommutatiivisia, osapuolten järjestys voidaan vaihtaa (kumpi on ulompi)
- $R \times S \equiv S \times R$
- $R \bowtie S \equiv S \bowtie R$
- Ristulo ja liitos ovat assosiativiset (suoritusjärjestys on vaihdettavissa)
- $R \times (S \times T) \equiv (R \times S) \times T$

5



## Kyselypuut ja ekvivalenssi

- Projektion ja valinnan järjestys on vaihdettavissa kunhan kaikki valinnassa esiintyvät sarakkeet ovat mukana projektiossa
 
$$\sigma_c(\pi_a(R)) \equiv \pi_a(\sigma_c(\dots(R)))$$
- Valinta ja ristitulo voidaan yhdistää liittokseksi, jos kyseessä on liittosehto
 
$$R \bowtie_c S \equiv \sigma_c(R \times S)$$
- Valinta voidaan siirtää tapahtuvaksi ennen ristituloa tai liittosta, mikäli se koskee vain näiden argumenttia
 
$$\sigma_c(R \times S) \equiv \sigma_c(R) \times S$$

6



## Kyselypuut ja ekvivalenssi

- Yleisesti valinta  $\sigma_c(R \times S)$  voidaan jakaa tauluja R ja S koskeviin osiin sekä liitosehtoon eli

$$\sigma_c(R \times S) \equiv \sigma_c(\sigma_{cR}(R) \times \sigma_{cS}(S))$$

- Projektio voidaan jakaa osiin jotka suoritetaan ennen ristituloa

$$\pi_a(R \times S) \equiv \pi_{a1}(R) \times \pi_{a2}(S)$$

- Projektio voidaan jakaa osiin, jotka suoritetaan ennen liitosta, jos liitosehdon sarakkeet säilyvät

$$\pi_a(R \bowtie S) \equiv \pi_{a1}(R) \bowtie \pi_{a2}(S)$$

7



## Heuristisia muokkaussääntöjä

- Konjunkttiiviset valinnat puretaan sarjaksi yksittäisiä valintoja
- Painetaan valinnat niin alas puussa kuin mahdollista
- Järjestellään lehtisolmut edulliseen järjestykseen
  - ensin mahdollisimman rajoittavat valinnat
  - = pieni rivijoukko odotettavissa tulokseksi
  - Yritetään välttää ristituloja
- Yhdistetään ristitulo ja seuraava valinta liitokseksi
- Yhdistetään operaatiot, jotka voidaan suorittaa yhdellä algoritmilla
- Vältellään välitulosten kirjoitusta

8



## Rajaavuus

- Rajaavuutta (selectivity) mitataan **rajaus-suhteella** (reduction factor):
- tuloksen rivimäärä / lähtöjoukon rivimäärä**
- Rajaavuuden arviointi ilman tilastotietoja on hankalaa
  - Kun testataan **avaita yhtäsuuruutta suhteessa vakioon** tuloksen koko on enintään 1 ja rajaussuhde **1/taulun koko** (jos taulun kokoa ei tiedetä, sen tilalla voidaan käyttää taululle tehdyn tilanvarauksen kokoa, tai jotain vakiota, jos tätäkään ei tiedetä.)
  - sarakkeen yhtäsuuruus verrattuna vakioon tuottaa tasaisen jakautuman oletuksella rajaussuhteen **1/sarakkeen arvojen lukumäärä**. (jos sarakkeella on hakemisto voi arvojen lukumäärä olla tiedossa, tietotyypin määrittäminen voi rajata arvojoukon esim. numeric(2) enintään 100 arvoa, järjestelmä voi käyttää oletusarvoa esim 1/10, jos sillä ei ole parempaa tietoa)

9



## Rajavuus

- Täsmällisemmän tiedon saanti rajaavuudesta edellyttää tilastotietoa:
  - taulun koko (riveinä, sivuina)
  - sarakkeen arvojen lukumäärä (riittää jos oletetaan tasainen jakautuma)
  - sarakkeen arvojakautuma (histogrammi) käyttökelpoinen, jos jakautuma on hyvin vino.

10