Tietokannan hallinta, kevät 2002 **Harjoitus 4 (22.-26.4.)**

Läsnäolo: 3 tehtävää

1. Arvioi dynaamiseen ja laajentuvaan hajautukseen perustuvien hakemistojen tilantarve ja levyhakujen määrä viikon 3 tehtävissä 1 ja 2 käsitellylle tiedostolle. (Tehtävän päätarkoitus on käydä läpi rakenteiden pääpiirteet.)

2. Vertaile luentojen sivuilla 8-9 esitettyjä kyselypuita seuraavilla monikkomäärillä: department 50, project 500, employee 5000. Tässä ei tarkastella asiaa relaatioiden toteutuksen eikä hakemistojen kannalta.
a) Mitkä ovat välitulosten koot sivun 8 puulle ja sen sellaisille muunnoksille, jotka saadaan laskemalla karteesiset tulot eri järjestyksessä?
b) Sivun 9 puuhun sisältyy kolme optimointitoimenpidettä: valinnan suoritus aikaisin, ja kummankin karteesisen tulon ja siihen sopivan valinnan korvaaminen liitoksella. Kuinka kukin näistä optimoinneista yksinään pienentää välituloksia?
c) Missä vaiheessa kyselyyn liittyvät projektiot tulisi suorittaa?

3. Osoita pienellä esimerkillä, että projektio ei ole vaihdannainen leikkauksen eikä erotuksen kanssa. (Asia ei ole vaikea, vaikka löytyy jopa oppikirjoja, joissa vaihdannaisuuden väitetään pätevän.)

4. Tarkastellaan yritystietokantaan kohdistettua kyselyä

```
select e.FNAME, e.LNAME, n.DEPENDENT_NAME
from EMPLOYEE e, DEPARTMENT d, DEPENDENT n
where e.SSN = n.ESSN and e.SSN = d.MGRSSN and d.DNAME = 'Research'
and n.RELATIONSHIP='SON'.
```

a) Mitä kysely tulostaa?
b) Anna kyselyä suoraan vastaava projektio-valinta-tulo-tyyppinen relaatiolauseke ja tätä vastaava kaavapuu. Assosioi karteesiset tulo-operaatiot tavalla, jonka arvelet tuottavan parhaan optimointituloksen.
c) Optimoi kysely käyttäen heuristiikkaa "suorita valinnat mahdollisimman aikaisin". Anna optimoitua kyselyä vastaava kaavapuu ja relaatiolauseke.
d) Voidaanko tehdä vielä muita optimointitoimenpiteitä?

5. a) Kuinka edellisen tehtävän kysely pitäisi laskea, kun relaatioon EMPLOYEE on hakemisto attribuutilla SSN ja relaatioon DEPARTMENT on hakemisto attribuutilla DNAME? Onko hakemistojen ominaisuuksilla merkitystä?
b) Voisiko kyselyn suoritusta tehostaa paremmin joillakin muilla hakemistoilla?
c) Onko hakemistojen ylläpito tässä kyselyssä esiintyville relaatioille viisasta – esimerkiksi suhteessa yritystietokannan muihin relaatioihin?

6. (**) Tarkastellaan kyselyä

```
select FNAME, LNAME, PNO, HOURS from EMPLOYEE, WORKS_ON
where ESSN = SSN and HOURS > 20;
```

Arvioi kyselyn suorituksessa tarvittavien levyhakujen määrää (ylä-ja alarajat), kun käytetään joko sisäkkäisten silmukoiden menetelmää tai hakemistoliitosta. EMPLOYEE-relaatiossa on 10000 kpl 100 merkin rivejä ja WORKS_ON-relaatiossa 30000 kpl 20 merkin rivejä. Jakson koko on 2 KB.