

CS-A1150 Tietokannat

Tentti 31.8.2023

Laskinten tai lisämateriaalin käyttö tentissä ei ole sallittua

1. (4 p) Kerro lyhyesti, miksi käytetään tietokantoja sen sijaan, että dataa säilytettäisiin vain tavallisissa tekstitiedostoissa. Mitä sellaisia etuja tietokannan hallintajärjestelmä tarjoaa, joita tekstitiedostot eivät suoraan tarjoa (ellei näitä ominaisuuksia ohjelmoi erikseen)? Suositeltu vastauksen pituus on 0,5-1 sivua.

2. a) (8 p) Tee seuraavan kuvauksen pohjalta UML-kaavio lääke- ja reseptitietokantaa varten. Tietokantaan tallennetaan tietoja lääkkeistä, potilaista, potilaille määrätystä lääkeresepteistä ja niiden toimittamisista (reseptillä määrättyjen lääkkeiden ostoista) sekä lääkäreistä. Käytä kursilla opetettua notaatiota ja merkitse avainattribuutit kursilla opetetulla tavalla. Tehtävän helpottamiseksi mallinnettavassa tietokannassa ei ole kaikkia niitä ominaisuuksia, joita todellisessa reseptitietokannassa olisi.

Potilaista on tietokantaan tallennettu henkilötunnus (yksikäsitteinen), nimi ja syntymäaika. Tietokantaan tallennetuilla resepteillä on yksikäsitteinen tunnus, reseptin määräyspäivämäärä, vanhenemispäivämäärä sekä määrätyn lääkkeen määrä ja yksikkö (esim. tablettien määrä kappaleina) ja merkijonona esitetty annosteluohje. Resepteistä tallennetaan myös, mikä lääke siinä on määrätty, kenelle potilaalle resepti on kirjoitettu ja kuka lääkäri on reseptin kirjoittanut. Yksinkertaisuuden vuoksi oletetaan, että yhdessä reseptissä voidaan määrätä vain yhtä lääkettä.

Tietokannassa on tieto erilaisista lääkkeistä ja niiden vaikuttavista aineista. Yhdessä lääkkeessä voi olla monta vaikuttavaa ainetta ja sama vaikuttava aine voi esiintyä useassa eri lääkkeessä. Lääkkeestä tiedetään sen nimi (yksikäsitteinen), pakkauskoko ja sen yksikkö, valmistajan nimi, mitä vaikuttavia lääkkeessä on ja kuinka paljon (määrä ja yksikkö) kutakin vaikuttavaa ainetta. Vaikuttavilla aineilla on latinankielinen nimi (yksikäsitteinen), suomenkielinen nimi ja ruotsinkielinen nimi.

Potilaan ei tarvitse ostaa koko reseptissä määrättyä määrää kerrallaan, vaan hän voi ostaa sen usealla eri ostokerralla (esimerkiksi jos reseptissä on määrätty yhteensä 500 tablettia, potilas voi ostaa vaikka kolmella eri kerralla jokaisella yhden 100 tabletin pakkauksen ja neljännellä kerralla kaksi 100 tabletin pakkausta). Tietokantaan on tallennettu jokaisesta ostokerrasta sen päivämäärä ja kellonaika sekä sillä kerralla ostetun lääkkeen määrä. Ostokerroilla ei ole erillistä tunnusta, vaan ne tunnistetaan oston ajankohdan ja ostoon liittyvän reseptin perusteella.

Kaikista lääkäreistä, joilla on Suomessa oikeus määrätä reseptejä, on tietokantaan tallennettu yksilöintitunnus (yksikäsitteinen), nimi ja tutkinto.

b) (2 p) Muuta a-kohdan UML-kaavio relaatiomalliin. Esitä relaatiokaaviot ja alleviivaa niistä avainattribuutit (eli PRIMARY KEY -määreellä merkittävät attribuutit).

3. Tehtävä liittyy seuraavaan tietokantakaavioon, joka kuvaa erään hotelliketjun asiakkaita, hotelleja, niiden huoneita ja hotelliin tehtyjä varauksia. Tehtävän yksinkertaistamiseksi oletetaan, että kukin varaus sisältää vain yhden huoneen ja varausta tehdessä määrätään aina samalla se huone, johon varaus hotellissa kohdistuu. Lisäksi oletetaan, että sama huone on aina saman hintainen.

Taulu `Hotels` sisältää tiedot ketjun hotelleista. Hotellit tunnistetaan yksikäsitteisen tunnuksen (attribuutti `ID`) avulla. Taulu `Customers` sisältää tiedot ketjun asiakkaista. Asiakkailla on yksikäsitteinen tunnus (attribuutti `CID`). Taulu voi sisältää myös sellaisia asiakkaita, jotka eivät ole tehneet vielä yhtään varausta, vaan joiden tiedot ovat järjestelmässä esimerkiksi siksi, että he ovat osallistuneet johonkin ketjun mainoskampanjaan.

Taulu `Rooms` sisältää tiedot ketjun hotellien huoneista. Attribuutti `number` kertoo huoneen numeron, `hotelID` sen hotellin tunnuksen, jossa huone on, `beds` huoneessa olevien vuoteiden lukumäärän ja `price` huoneen hinnan vuorokautta kohti.

Taulu `Reservations` sisältää tiedon hotelliketjun huonevarauksista. Varauksilla on yksikäsitteinen varaustunnus `RID`. Taulun attribuutit sisältävät tiedon varauksen tehneestä asiakkaasta, varatun huoneen numeron ja missä hotellissa se sijaitsee, varauksen alkamispäivämäärän ja sen, kuinka moneksi peräkkäiseksi yöksi varaus on tehty. Varaukseen kuuluu aina vain yksi huone.

Huoneen numero, vuoteiden lukumäärä ja varaukseen kuuluvien öiden lukumäärä ovat kokonaislukuja ja huoneen hinta desimaaliluku. Muiden attribuuttien arvot ovat merkkijonoja. Voit olettaa, että tauluissa olevien monikoiden (rivien) attribuuteilla ei ole `NULL`-arvoja.

Tietokantakaavio:

`Hotels(ID, name, city, address, country)`

`Customers(CID, name, email, phone)`

`Rooms(number, hotelID, beds, price)`

`Reservations(RID, customerID, hotelID, roomNo, date, numberofnights)`

Kirjoita seuraavat kyselyt SQL-kielillä:

- a) (2 p) Niiden hotellien ID:t ja nimet, joissa on vähintään yksi huone, jonka hinta on alle 60 euroa.
- b) (2 p) Niiden asiakkaiden tunnuksot ja nimet, jotka ovat tehneet vähintään kerran varauksen, jonka kokonaisarvo (huoneen hinta kertaa öiden lukumäärä) on yli 1000 euroa.
- c) (2 p) Niiden hotellien nimet ja tunnuksot, joissa on vähintään yksi huone, jossa on kolme sänkyä, mutta ei yhtään huonetta, jossa olisi neljä sänkyä tai enemmän.
- d) (2 p) Haetaan ne ketjun Suomessa sijaitsevat hotellit, joihin mahtuu yhteensä vähintään 100 vierasta (vuoteiden yhteislukumäärä hotellin kaikissa huoneissa). Kyselyn pitää tuottaa näistä hotelleista hotellin ID, nimi ja huoneiden hintojen keskiarvo. Vuodepaikkojen lukumäärä ja huonehintojen keskiarvo lasketaan siis erikseen kullekin hotellille.

Kirjoita seuraavat kyselyt relaatioalgebran lausekkeina:

e) (2 p) Niiden asiakkaiden tunnukset ja nimet, jotka ovat varanneet huoneen Huippuloma-nimisessä hotellissa.

f) (2 p) Niiden hotellien ID:t ja nimet, joissa kaikilla saman hotellin huoneilla on sama hinta.

4. Tarkastellaan relaatiota $R(A, B, C, D, E)$, jossa on voimassa funktionaaliset riippuvuudet $A \rightarrow B$, $B \rightarrow C$ ja $B \rightarrow E$.

a) (1 p) Perustele, miksi relaatio ei ole Boyce-Codd-normaali muodossa (BCNF:ssä).

b) (6 p) Osita relaatio Boyce-Codd-normaali muotoon käyttämällä kurssilla (ja oppikirjassa) esitettyä algoritmia. Perustele lyhyesti jokainen muodostamasi uusi relaatio. Jatka osittamista niin pitkälle, että jäljellä on vain BCNF:ssä olevia relaatioita. Perustele, miksi lopulliset relaatiot ovat BCNF:ssä.

5. (7 p) Kerro lyhyesti, mitkä ovat transakzioilta (engl. transaction) vaadittavat ominaisuudet (ns. ACID-ominaisuudet) tietokannan hallintajärjestelmässä. Esitä jokaisesta ominaisuudesta jokin virhetilanne, joka voisi tapahtua tehtävässä 3 annetussa hotellitietokannassa, jos tämä ominaisuus ei ole voimassa. C-kirjainta vastaavan ominaisuuden (suomeksi vastaavan ominaisuuden nimi alkaa E-kirjaimella) osalta esitä kolme erityyppistä virhetilannetta. Muiden kolmen ominaisuuden osalta riittää jokaisesta yhden esimerkin esittäminen.