

Proeftentamen Databases 1, 2M400, 6 juni 2003.

De uitwerkingen van de opgaven dienen duidelijk en bondig geformuleerd te worden, en overzichtelijk opgeschreven. Vermeld de naam van je colstructeur, want hij of zij zal het proeftentamen nakijken.

Dit is een “open boek” proeftentamen. Je mag gebruik maken van boeken en aantekeningen, op papier of op computer. Alle niet door jezelf geschreven materiaal dient te dateren van vóór de aanvang van dit proeftentamen. Het gebruik van communicatie-apparatuur is niet toegestaan. (Geen mobiele telefoons, alleen laptops zonder netwerk.)

In de query opgaven wordt een (deeltje van een) universiteitsbibliotheek-database gebruikt met de volgende tabellen en attributen:

boek : { ISBN, titel, uitgever, jaar }
auteur : { ISBN, voorletters, naam }
exemplaar : { barcode, ISBN, faculteit, exjaar, aanwezig }
reservering : { naam, faculteit, ISBN, datum, geannuleerd }
uitlening : { naam, faculteit, barcode, van, tot }

Korte beschrijving:

Elk boek heeft een uniek ISBN nummer. Het heeft een titel, uitgever, jaar van uitgifte en een aantal auteurs waarvan we de voorletters (samen als 1 string) en de familienaam bijhouden. (Verschillende edities van eenzelfde boek hebben een verschillend ISBN nummer en zijn dus voor ons verschillende boeken.)

De universiteit heeft van elk boek 1 of meer exemplaren. Elk exemplaar is in een bepaald jaar (exjaar) aangekocht, heeft een sticker met unieke barcode en is toegewezen aan een faculteit. (Boeken kunnen worden uitgeleend aan mensen van andere faculteiten.) Een exemplaar kan afwezig zijn omdat het nog niet geleverd is, wordt hersteld, of omdat het is uitgeleend.

Personen worden geïdentificeerd door hun (familie)naam en faculteit. Ze kunnen een boek reserveren op een bepaalde datum. (Het type van “datum” is zodanig dat data die niet NULL zijn met elkaar vergeleken kunnen worden, bijvoorbeeld met $<$ en \leq .) Personen kunnen een reservering op elk willekeurig moment weer annuleren. De datum blijft dan die van de reservering, dus de datum waarop de reservering is geannuleerd wordt niet bijgehouden. Ze kunnen een exemplaar van een boek lenen op een bepaalde “van” datum. Een eventuele reservering van dat boek (op hun naam) wordt dan automatisch geannuleerd. De “tot” datum blijft NULL tot het boek is teruggebracht. De “tot” datum is minstens 1 dag later dan de “van” datum. Wanneer een boek is teruggebracht wordt de “uitlening” in de database bewaard en wanneer een reservering wordt geannuleerd (bijvoorbeeld omdat het boek wordt geleend maar het kan ook om andere redenen zijn) wordt die reservering ook bewaard.

De bibliotheek bestaat al een hele tijd. Bijgevolg heeft elke faculteit wel wat (exemplaren van) boeken, heeft er uit elke faculteit wel iemand een boek gereserveerd en iemand een exemplaar geleend, en is er van elke faculteit wel eens een boek gereserveerd en wel eens een (exemplaar van) een boek geleend.

Let op: De attributen “naam” en “faculteit” worden in deze database in twee verschillende betekenissen gebruikt. Zorg dat je die niet per ongeluk door elkaar gebruikt of in een join mee neemt. Let ook op voor de verschillen tussen “boeken” en “exemplaren”. In een vraag schrijven we misschien wel eens dat een “boek” wordt geleend, maar we bedoelen dan dat een “exemplaar” van dat “boek” wordt geleend.

1. Bestudeer grondig de beschrijving van de bibliotheek-database.
Beschrijf dan vier van de volgende vijf vragen in de relationele algebra, de tupel calculus of SQL naar keuze. Je moet wel tenminste 1 keer de algebra, 1 keer de calculus en 1 keer SQL gebruiken, en precies 1 van de formalismen per vraag. Als je een aggregatie wil gebruiken moet je dat met SQL doen.
 - (a) Geef de naam en voorletters van de auteurs van wie er volgens de database een exemplaar van een boek aanwezig is terwijl datzelfde exemplaar ook uitgeleend is.
 - (b) Geef de naam en voorletters van de auteurs van wie er nog nooit een boek gereserveerd is.
 - (c) Geef de naam en faculteit van personen die reeds van elke faculteit een (exemplaar van een) boek hebben geleend.
 - (d) Geef de naam en voorletters van de auteurs van wie alle exemplaren van al hun boeken ooit tegelijk uitgeleend waren.
 - (e) Geef de naam en faculteit van personen die op dit moment meer dan 10 boeken hebben geleend (en dus nog niet teruggebracht).
2. Leg in normaal klinkend Nederlands uit wat er in de volgende query wordt gevraagd:

```
SELECT DISTINCT b.titel
FROM boek as b, exemplaar as e
WHERE b.ISBN = e.ISBN
      AND e.barcode IN
      ( SELECT u1.barcode
        FROM uitlening as u1, uitlening as u2
        WHERE u1.barcode = u2.barcode
              AND u2.van <= u1.tot
              AND (u1.van < u2.van OR u2.van < u1.tot)
              AND (u1.tot <= u2.tot OR u2.tot IS NULL) )
```

3. Toon aan (bewijs) dat elk relatie schema $R = (\{A, B\}, \Delta, dom, SC)$ (dus met twee attributen) in BCNF is als $\emptyset \rightarrow A \notin SC$ en $\emptyset \rightarrow B \notin SC$.
4. Het relatieschema "O" (voor "Onderwijs") heeft attributen "docent", "college", "dictaat" en "zaal". Een tupel betekent dat een docent een college verzorgt met een bepaald dictaat en in een bepaalde zaal. "O" voldoet aan de volgende beperking: Alle docenten die eenzelfde college verzorgen gebruiken hierbij dezelfde (set van) dictaten. Wat voor beperking is dit, **en waarom?** (Let op, zonder uitleg geen punten.)
 - (a) functionele afhankelijkheid
 - (b) multivalued afhankelijkheid
 - (c) join afhankelijkheid
 - (d) iets anders (een soort die niet op college behandeld is)
5. Als je weet dat de Armstrong regels F1, F2 en F3 samen "volledig" zijn, bewijs dan dat F1 en F6 samen volledig zijn. (F1 = reflexivity, F6 = pseudotransitivity).

Waardering: vraag 1: 1 punt per query, vraag 2: 1 punt, vraag 3: 2 punten, vraag 4: 1 punt, vraag 5: 2 punten.

Suggestie voor verdeling van 90 minuten proeftentamentijd: vraag 1: 30 minuten (2 vragen van 5 minuten, 2 vragen van 10 minuten, 1 vraag overslaan), vraag 2: 10 minuten, vraag 3: 10 minuten, vraag 4: 5 minuten, vraag 5: 15 minuten, reserve aan het einde: 20 minuten.

Uitwerkingen

1. Er zijn natuurlijk veel mogelijkheden om een keuze te maken voor de algebra, tupel calculus en SQL. We werken meestal maar 1 keuze uit.

- (a) We lossen de eerste vraag bijvoorbeeld op in SQL omdat we in de relationele algebra en in de tupel calculus niet zo goed kunnen omgaan met een NULL waarde.

```
SELECT DISTINCT a.naam, a.voorletters
FROM auteur AS a, exemplaar AS e, uitlening AS u
WHERE a.ISBN = e.ISBN AND e.barcode = u.barcode
      AND e.aanwezig = true AND u.tot IS NULL
```

Merk op dat als je deze vraag zou willen oplossen in de algebra het attribuut “faculteit” in “exemplaar” en “uitlening” een andere betekenis heeft en dus niet in een join mag worden gematcht.

- (b) De tweede vraag kan makkelijk in tupel calculus:

$$\{t \mid \exists a \in \text{auteur} (t[\text{naam}] = a[\text{naam}] \wedge t[\text{voorletters}] = a[\text{voorletters}] \wedge \\ \forall r \in \text{reservering} (\forall aa \in \text{auteur} (r[\text{ISBN}] = aa[\text{ISBN}] \Rightarrow \\ (aa[\text{naam}] \neq a[\text{naam}] \vee aa[\text{voorletters}] \neq a[\text{voorletters}])))\})\}$$

Bij deze vraag is het opletten dat het gaat om auteurs van wie nog geen enkel boek gereserveerd is. In de algebra gaat dat behalve met de bovenstaande uitwerking ook gemakkelijk met een verschil. Er is altijd twee keer “auteur” nodig omdat je anders kijkt naar het isbn van 1 boek en niet van alle boeken van de auteur.

- (c) De derde vraag gaat over een vergelijking met alle faculteiten wat een statische verzameling is, dus dringt zich de relationele algebra op met de division operator. Het is wel even opletten want faculteit komt in twee betekenissen voor. Laat f de bijectie zijn op de attributen van exemplaar die faculteit omnoemt naar efaculteit en de overige attributen gelijk laat. Dan wordt de query:

$$\Pi_{\{\text{naam}, \text{faculteit}, \text{efaculteit}\}}(\text{uitlening} \bowtie \rho_f(\text{exemplaar})) \div \Pi_{\{\text{efaculteit}\}}(\rho_f(\text{exemplaar}))$$

Hoewel een “division” query ook in tupel calculus of SQL kán worden uitgewerkt is hierin veel sneller een foutje gemaakt.

- (d) De vierde vraag was bedoeld om over te slaan, maar dat wil niet zeggen dat ze onmogelijk te doen was.

We moeten alle auteurs vinden voor wie een dag bestaat toen alle (exemplaren van) hun boeken uitgeleend waren. We nemen aan dat exemplaren op de begindatum van de uitlening in uitlening zijn en op de einddatum net niet meer. (Dus als een exemplaar om 10 uur teruggebracht is en een ander om 14 uur uitgeleend wordt, zijn de uitleningen niet tegelijk want de eerste uitlening wordt als afgelopen beschouwd op 0.00 uur.)

Een observatie: Als voor auteur a een periode bestaat toen alle (exemplaren van) zijn boeken uitgeleend waren, dan bestaat er (ten minste) een uitlening van zijn boek die op de begindatum van die periode begint. Alle (exemplaren van) zijn boeken zijn op die dag uitgeleend. In de algebra is dit echt moeilijk uit te drukken maar in de tupel calculus en in SQL valt het nog wel mee:

$$\{t \mid \exists a \in \text{auteur} (t[\text{naam}] = a[\text{naam}] \wedge t[\text{voorletters}] = a[\text{voorletters}] \wedge \\ \exists u \in \text{uitlening} (\exists e \in \text{exemplaar} (e[\text{barcode}] = u[\text{barcode}] \wedge e[\text{ISBN}] = a[\text{ISBN}] \wedge \\ \forall e1 \in \text{exemplaar} ((\exists a1 \in \text{auteur} (a1[\text{naam}] = a[\text{naam}] \wedge a1[\text{voorletters}] = a[\text{voorletters}] \\ \wedge e1[\text{ISBN}] = a1[\text{ISBN}])) \Rightarrow \\ \exists u1 \in \text{uitlening} (u1[\text{barcode}] = e1[\text{barcode}] \wedge u1[\text{van}] \leq u[\text{van}] \wedge \\ (u1[\text{tot}] > u[\text{van}] \vee u1[\text{tot}] = \text{NULL}))))))\}$$

De vertaling naar SQL:

```
SELECT a.naam, a.voorletters
FROM  auteur AS a, uitlening AS u, exemplaar AS e
WHERE e.barcode = u.barcode AND e.ISBN = a.ISBN AND
      NOT EXISTS
      (SELECT *
       FROM  exemplaar AS e1, auteur AS a1
       WHERE a1.naam = a.naam AND a1.voorletters = a.voorletters AND
             e1.ISBN = a1.ISBN AND
             NOT EXISTS
             (SELECT *
              FROM  uitlening AS u1
              WHERE u1.barcode = e1.barcode AND u1.van <= u.van
                    AND (u1.tot > u.van OR u1.tot IS NULL)))
```

(e) De vijfde vraag is een aggregaat-query dus die lossen we op in SQL:

```
SELECT DISTINCT u.naam, u.faculiteit
FROM  uitlening AS u
WHERE u.tot IS NULL
GROUP BY u.naam, u.faculiteit
HAVING COUNT(barcode) > 10
```

Dit kan ook zonder aggregatie en dan mag het ook in de algebra of tupel calculus.

- De wat ingewikkeld uitziende expressie aan het eind probeert aan te geven dat de periodes van uitleningen u_1 en u_2 overlappen maar niet identiek zijn. De query betekent dan: “Geef de titels van exemplaren van boeken die volgens de database ooit tweemaal tegelijk uitgeleend waren.”
Helaas is er een foutje in die expressie, waardoor de query eigenlijk gewoon alle titels van boeken die ooit uitgeleend en teruggebracht zijn oplevert. Zowel het echt correcte als het bedoelde antwoord worden hier goed gerekend.
- Als $\emptyset \rightarrow A$ en $\emptyset \rightarrow B$ niet gelden dan zijn de enige mogelijke niet-triviale functionele afhankelijkheden $A \rightarrow B$, $A \rightarrow AB$, $B \rightarrow A$ en $B \rightarrow AB$ en dat zijn allemaal sleutelafhankelijkheden. Aangezien er geen niet-triviale functionele afhankelijkheden zijn die geen sleutelafhankelijkheid zijn is het schema in BCNF.
Let op dat we niet beweren dat $A \rightarrow B$, $A \rightarrow AB$, etc. ook gelden in het schema. We weten niet welke afhankelijkheden er gelden, maar als ze gelden zijn het sleutelafhankelijkheden.
- Het juiste antwoord is *vanzelfsprekend* d:
Dat de beperking geen functionele afhankelijkheid komt omdat een college een *verzameling* dictaten (en docenten) bepaalt, dus niet één enkel dictaat (of één enkele docent), wat voor een functionele afhankelijkheid zou moeten.
De beperking lijkt op een multivalued afhankelijkheid (of dus ook een binaire join afhankelijkheid) maar ze is het niet want een multivalued afhankelijkheid of join afhankelijkheid zegt iets over alle attributen en deze beperking zegt maar iets over drie van de vier attributen. Het is dus een andere soort beperking (antwoord d). Ter informatie: het is een “embedded multivalued afhankelijkheid”, een soort beperking die we niet op college behandeld hebben.
- Om de volledigheid van F1 en F6 te bewijzen moeten we aantonen dat we door gebruik te maken van (alleen maar) F1 en F6 de regels F1, F2 en F3 kunnen afleiden (F1 is triviaal):
F2: Definieer δ als $\beta\gamma$.
Als $\alpha \rightarrow \beta$ gegeven is, en $\beta\gamma \rightarrow \delta$ ($= \beta\gamma$) geldt wegens F1, dan geldt $\alpha\gamma \rightarrow \delta$ wegens F6 en dat is dus $\alpha\gamma \rightarrow \beta\gamma$ zoals F2 het wil.
F3: Dit volgt direct uit F6 door $\gamma = \emptyset$ te stellen.