

Naam en voorletters:

Ident.nr.:

TECHNISCHE UNIVERSITEIT EINDHOVEN

---

Faculteit Wiskunde en Informatica

Proeftentamen Databases 1, 2M400, 9 en 11 juni 2004.

Alle uitwerkingen van de opgaven **moeten** worden ingevuld in de daarvoor bestemde vrije ruimte bij de tentamenvragen. **Alleen wat op deze formulieren is ingevuld wordt nagekeken. Andere papieren worden meteen vernietigd. Ze worden in geen geval nagekeken.** De ruimte die is voorzien is in elk geval voldoende voor het correcte antwoord, in een normaal leesbare letter-grootte. Alles wat onvoldoende leesbaar is wordt fout gerekend! Elk antwoord moet staan in de ruimte bij de vraag, tenzij verwezen is naar de laatste (witte) pagina.

---

Dit is een “open boek” tentamen. Je mag gebruik maken van boeken en aantekeningen, op papier of op de computer. Alle niet door jezelf geschreven materiaal dient te dateren van vóór de aanvang van dit tentamen. Het gebruik van communicatieapparatuur is niet toegestaan. (Geen mobiele telefoons, wel laptops zonder netwerk, voor het lezen van elektronisch materiaal.)

---

Hier is een recept voor het met succes afronden van dit proef-tentamen:

1. Bestudeer eerst alle vragen, en stel een lijstje op met de volgorde waarin je de vragen wil beantwoorden.
2. Begin met het beantwoorden van de vragen waarvan je zelf denkt dat je ze het beste kan. Werk zorgvuldig. Het is niet erg om wegens tijdgebrek aan het eind een vraag open te moeten laten.
3. Denk bij elke vraag eerst goed na over het “type” van de vraag: welk deel van het collegemateriaal is van belang bij het beantwoorden van de vraag? Bij een query: wat voor soort query is het? Bij afleidingsregels: welke “richting” moet ik bewijzen? etc.
4. Werk elke vraag eerst op een kladblaadje uit en schrijf het antwoord dan over op het invulblad. **Wacht hiermee niet tot het einde want losse bijgevoegde papieren worden niet nagekeken! Sterker: ze worden door de surveillant geweigerd, en als hij/zij ze toch zou aannemen worden ze door de corrector meteen vernietigd!** Je mag bij een vraag verwijzen naar de laatste pagina, maar als die vol is houdt het op!
5. Als je nog 15 a 20 minuten hebt, begin dan niet meer aan een nieuwe vraag (die je niet ligt en daarom hebt uitgesteld tot het einde) maar kijk de vragen na die je al hebt beantwoord. Bij queries: vertaal ze zonder naar de opgave te kijken eens terug, en kijk dan of wat uit die vertaling komt ook de opgave was. Bij de afhankelijkheden of decompositie, ga na of je antwoord wel het antwoord is op de vraag die is gesteld: heb je de juiste “richting” bewezen? Heb je geen veronderstellingen gemaakt die de algemeenheid van je uitwerking schaden, voldoet bij een decompositie je resultaat wel aan de gevraagde normaalvorm, etc.
6. Verifieer bij elke formule (in eender welke taal) of het aantal haakjes en accolades klopt, en of de formule helemaal voldoet aan de syntax van de taal waarin ze geschreven is. Formules of queries met syntaxfouten worden meteen helemaal fout gerekend, ongeacht wat de formule of query probeert uit te drukken.
7. Verifieer dat je op elke pagina je naam en identiteitsnummer hebt geschreven.

Naam en voorletters:

Ident.nr.:

## De bibliotheek-database

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 achternaam bijhouden. Verschillende edities van eenzelfde boek hebben een verschillend ISBN nummer en zijn dus voor ons verschillende boeken. De database bevat alle auteurs van in de database aanwezige boeken, en geen auteurs die niet (met ISBN nummer) overeenkomen met een boek..

De universiteit heeft van elk boek 1 of meer exemplaren. Elk exemplaar is in een bepaald jaar (exjaar) aangekocht, heeft een sticker met een 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 achternaam 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 ≤.) 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. (Het "geannuleerd" attribuut heeft altijd een waarde "ja" of "nee".)

Ze kunnen een exemplaar van een boek lenen op een bepaalde "van" datum. Ze zijn dan "lener". 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 eens 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 zeker op de volgende "valstrikken":

- 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" (eender welk!) van dat "boek" wordt geleend. Twee uitleningen van eenzelfde boek kunnen uitleningen van verschillende exemplaren zijn. Zorg dat je altijd goed in de gaten hebt of er een "boek" als generiek object wordt bedoeld of een "exemplaar" als concreet item dat kan worden geleend.
- Wie alleen nog maar een boek gereserveerd heeft maar nog nooit iets heeft geleend is nog geen "lener". Een "lener" is iemand die al eens een boek heeft geleend.

Naam en voorletters:

Ident.nr.:

**Algemene informatie** (heeft geen enkele invloed op het cijfer):

Mijn colstructeur was: (aankruisen of omcirkelen svp)

De Bra                      Aerts                      Voorhoeve                      n.v.t.

Mijn lab-instructeur was: (aankruisen of omcirkelen svp)

Frasincar                      Vdovjak                      Thiran                      n.v.t.

n.v.t. betekent: ik volgde weinig of geen colstructies of labsessies

1. Gegeven is het volgende relatie schema voor een vervoerbedrijf dat auto's, bestelwagens (met of zonder grote aanhanger), vrachtwagens en bussen verhuurt met chauffeur:

**persnr, naam, rijbewijs, kenteken, voertuigtype, vereist-rijbewijs**

De chauffeurs hebben een uniek nummer (persnr) en een naam en beschikken over een type rijbewijs (B, B+E, C, C+E, D, etc.). De voertuigen hebben een uniek kenteken, een type (auto, bestelwagen, bestelwagen met aanhanger, vrachtwagen, bus), en voor elk voertuig is een bepaald rijbewijs vereist (vereist-rijbewijs). Voor sommige kleine vrachtwagens kan rijbewijs B volstaan terwijl voor grotere het rijbewijs C vereist is. Voor de eenvoud wordt bij elke chauffeur niet alleen zijn rijbewijstype bijgehouden maar ook alle "mindere" types. (Dus wie bijvoorbeeld een rijbewijs D heeft heeft ook C en B en wie C+E heeft heeft ook C en B+E en B). Elk voertuig heeft 1 type en 1 rijbewijs-vereiste. Wanneer een persoon aan een kenteken is gekoppeld betekent dit dat deze persoon met dat voertuig mag rijden van het bedrijf. Deze toewijzingen moeten voldoen aan de beperking dat een persoon alleen mag worden toegewezen aan een voertuig waarvoor hij een geldig rijbewijs heeft.

Geef aan welke beperkingen/constraints in deze beschrijving staan. Geef de woorden en de formule.

voorbeeld:      de chauffeurs hebben een uniek nummer en een naam  
                    { persnr } → { naam }

voertuigen hebben een uniek kenteken, type en vereist-rijbewijs  
herhaald in: elk voertuig heeft 1 type en 1 rijbewijs-vereiste

{ kenteken } → { voertuigtype, vereist-rijbewijs }

een persoon mag alleen worden toegewezen aan een voertuig waarvoor hij een geldig rijbewijs heeft. dit is niet een constraint die we voor decompositie zullen gebruiken maar het is een inclusion dependency: (even het symbool  $\subseteq$  misbruikt)

{ persnr, vereist-rijbewijs }  $\subseteq$  { persnr, rijbewijs }

Foute constraint: { persnr } → { rijbewijs }

Voor de eenvoud wordt bij elke chauffeur niet alleen zijn rijbewijstype bijgehouden maar ook alle "mindere" types. Bij elke chauffeur kunnen dus meer rijbewijzen horen

Naam en voorletters:

Ident.nr.:

Sommigen geven  $\{\text{persnr}\} \rightarrow \{\text{rijbewijs}\}$  als constraint.

Dit zou correct kunnen zijn maar staat niet zo expliciet in de tekst. Opdat dit waar is moet namelijk de verzameling rijbewijzen die we bij een persoon bijhouden echt onafhankelijk zijn van de overige informatie in de hele tabel. We rekenen het waarnemen van deze constraint goed. Maar we kunnen er later niets mee doen omdat we een BCNF decompositie gaan maken en mvd's doen daarin niet mee.

Ook fout:  $\{\text{persnr}\} \rightarrow \{\text{kenteken}\}$

Een chauffeur is niet verbonden met slechts 1 voertuig. Hij kan aan vele voertuigen worden toegewezen (mits hij het goede rijbewijs heeft voor die voertuigen).

Ook fout:  $\{\text{voertuigtype}\} \rightarrow \{\text{vereist-rijbewijs}\}$

Het bedrijf gebruikt het type "vrachtwagen" voor lichte en zware vrachtwagens, waarvoor B dan wel C vereist is. Zie:

"Voor sommige kleine vrachtwagens kan rijbewijs B volstaan terwijl voor grotere het rijbewijs C vereist is."

Vervang dit relatieschema door een database schema (met verschillende relaties) waarin alle eigenschappen uit bovenstaande beschrijving kunnen worden weergegeven en dat in BCNF is. (Dus, lossless-join decompositie in BCNF maken.)

D.m.v.  $\{\text{persnr}\} \rightarrow \{\text{naam}\}$  kunnen we decomponeren in:

R1 ( persnr, naam )

R2 ( persnr, rijbewijs, kenteken, voertuigtype, vereist-rijbewijs )

R1 is dan in BCNF

D.m.v.  $\{\text{kenteken}\} \rightarrow \{\text{voertuigtype, vereist-rijbewijs}\}$  kunnen we R2 decomponeren in:

R21 ( kenteken, voertuigtype, vereist-rijbewijs )

R22 ( kenteken, persnr, rijbewijs )

(De enige kandidaatsleutel is steeds onderstreept)

R21 is in BCNF omdat er alleen nog een sleutelfafhankelijkheid geldt. (Er zijn geen andere afhankelijkheden tussen de attributen.

R22 is in BCNF omdat er geen enkele niet-triviale functionele afhankelijkheid geldt.

De hele decompositie wordt dan: R1, R21, R22.

Je zou R22 nog kunnen opsplitsen met behulp van  $\{\text{persnr}\} \rightarrow \{\text{rijbewijs}\}$  en dan krijg je R221 ( kenteken, persnr ) en R222 ( persnr, rijbewijs ) en deze decompositie is dan in 4NF, wat meer is dan er werd gevraagd maar wat wel logisch is.

Naam en voorletters:

Ident.nr.:

Geef aan wat je in dit schema moet veranderen wanneer het bedrijf de volgende twee beperkingen invoert:

- a. Elke chauffeur mag nog maar met voertuigen rijden waarvoor eenzelfde type rijbewijs vereist is. (Dus een chauffeur rijdt bijvoorbeeld met alleen maar voertuigen waarvoor rijbewijs C+E vereist is.)
- b. Elke chauffeur wordt toegewezen aan alle voertuigen met de rijbewijs-vereiste die aan de chauffeur is gekoppeld.

Je moet ervoor zorgen dat het nieuwe schema opnieuw in BCNF is, en je moet elke verandering motiveren. (Er volgt puntenaftrek wanneer de motivatie ontbreekt.)

Beperking a betekent  $\{ \text{persnr} \} \rightarrow \{ \text{vereist-rijbewijs} \}$

Deze functionele afhankelijkheid heeft geen invloed op de afhankelijkheden die gelden in R1, R21 en R22 en ze kan in geen van deze relaties worden voorgesteld.

Bijgevolg is er geen bijkomende decompositie nodig voor deze constraint.

Beperking b betekent  $\{ \text{vereist-rijbewijs} \} \twoheadrightarrow \{ \text{kenteken} \}$

(of  $\{ \text{vereist-rijbewijs} \} \twoheadrightarrow \{ \text{persnr} \}$ ) in de relatie

$\Pi_{\text{persnr}, \text{kenteken}, \text{vereist-rijbewijs}}(\mathbf{R})$

Ook deze afhankelijkheid kan in R1, R21 en R22 niet worden voorgesteld dus verandert deze ook niets aan de vereiste decompositie.

Bijgevolg is geen schemawijziging nodig om BCNF te bewaren. (Maar de decompositie is helaas niet meer dependency-preserving.)

Naam en voorletters:

Ident.nr.:

2. Druk de volgende vragen over de bibliotheek uit in één van de bestudeerde querytalen (relationele algebra, tupel calculus of SQL) naar keuze. Hint: sommige vragen zijn in sommige talen gemakkelijker dan in andere. Je mag het jezelf dus gemakkelijk maken door een verstandige keuze van query taal. Je **moet** bij elke vraag een andere querytaal gebruiken, dus 1 vraag in de relationele algebra, 1 in tupel calculus en 1 in SQL. Heel opgave 2 wordt fout gerekend als je tweemaal eenzelfde querytaal gebruikt.
- a. **Geef de titels van boeken waarvan de bibliotheek meer dan 1 exemplaar heeft aangekocht in het jaar van uitgifte.**

Welk formalisme (query taal) kies je, en waarom?

Vraag b wil ik in RA doen met een division. Dat laat TC en SQL over voor a en c. Laten we maar een keer SQL nemen voor deze vraag. Dan kunnen we tellen.

Uitwerking:

```
select b.titel
from boek as b
where 1 <
( select count(*)
  from exemplaar as e
  where e.exjaar = b.jaar and e.ISBN = b.ISBN )
```

Zitten er in deze opgave 1 of meer addertjes onder het gras? Zoja welke en wat heb je gedaan om het addertje (of de addertjes) te vermijden?

Als je deze vraag zonder “count” oplost moet je twee exemplaren vinden en garanderen dat ze een verschillende barcode hebben. Als je die ongelijkheid vergeet is het fout.

- b. **Geef de titel van de boeken die al werden gereserveerd door mensen van alle faculteiten.**

Welk formalisme kies je, en waarom? Herken je in deze vraag een “type” dat je bij de keuze van het formalisme of de aard van de uitwerking gebruikt?

RA omwille van division.

Uitwerking:

$$\Pi_{\text{titel}}(\text{boek} \bowtie (\Pi_{\text{ISBN, faculteit}}(\text{reservering}) \div \Pi_{\text{faculteit}}(\text{reservering})))$$

Zitten er in deze opgave 1 of meer valkuilen? Zoja welke en wat heb je gedaan om niet in die valkuil(en) te trappen?

Je moet zien dat dit in RA het makkelijkste is, en opletten dat boeken worden geïdentificeerd door ISBN, niet door titel.

Naam en voorletters:

Ident.nr.:

Het volgende is dus fout:

$$\Pi_{\text{titel, faculteit}}(\text{boek} \triangleright \triangleleft \text{reservering}) \div \Pi_{\text{faculteit}}(\text{reservering})$$

Ook fout is het om eerst te delen en dan te projecteren:

$$\Pi_{\text{titel}}(\text{boek} \triangleright \triangleleft (\text{reservering} \div \Pi_{\text{faculteit}}(\text{reservering})))$$

Want je deelt dan de hele reservering tabel door de faculteiten en krijgt dan de boeken die al door mensen met eenzelfde naam, maar uit alle faculteiten zijn gereserveerd op eenzelfde datum. Dus wanneer er in elke faculteit iemand is die "Jansen" heet, en al die Jansens hebben reeds eenzelfde boek gereserveerd, op dezelfde datum, en wanneer de status van die reserveringen gelijk is (allemaal geannuleerd of geen enkele geannuleerd), dan komt de titel van dat boek voor in het antwoord.

- c. **Geef de titels van boeken die nog nooit door eenzelfde persoon (tenminste) twee keer zijn uitgeleend.**

Welk formalisme kies je, en waarom? Herken je in deze vraag een "type" dat je bij de keuze van het formalisme of de aard van de uitwerking gebruikt?

TC blijft over.

Uitwerking:

$$\{ t \mid \exists b \in \text{boek} ( t[\text{titel}] = b[\text{titel}] \wedge \forall u1 \in \text{uitlening} ( \forall u2 \in \text{uitlening} ( \exists e1 \in \text{exemplaar} ( \exists e2 \in \text{exemplaar} ( ( u1[\text{barcode}] = e1[\text{barcode}] \wedge u2[\text{barcode}] = e2[\text{barcode}] \wedge u1[\text{naam}] = u2[\text{naam}] \wedge u1[\text{faculteit}] = u2[\text{faculteit}] \wedge e1[\text{isbn}] = e2[\text{isbn}] \wedge e1[\text{isbn}] = b[\text{isbn}] \Rightarrow ( u1[\text{van}] = u2[\text{van}] \wedge u1[\text{barcode}] = u2[\text{barcode}] ) ) ) ) ) ) ) \}$$

Zitten er in deze opgave 1 of meer instinkers? Zo ja welke en wat heb je gedaan om er niet in te stinken?

Let op! Het gaat om 2 uitleningen van hetzelfde boek, niet hetzelfde exemplaar. Let ook goed op met de haakjes. Let er ook op dat personen worden geïdentificeerd door naam en faculteit, niet door de naam alleen.

Let ook op: de tupel calculus laat niet toe om twee kwantoren te combineren. Bij  $\exists e2, e3 \in \text{exemplaar}$  is er nog geen probleem, maar wat betekent  $\neg \exists e2, e3 \in \text{exemplaar}$ ? Is dit  $\neg \exists e2 \in \text{exemplaar} ( \exists e3 \in \text{exemplaar}$  of is dit  $\neg \exists e2 \in \text{exemplaar} ( \neg \exists e3 \in \text{exemplaar} \dots$

Ook opletten:

$\{ t \mid \exists b \in \text{boek} ( t[\text{titel}] = b[\text{titel}] \Rightarrow \dots \}$  maar de implicatie betekent:

$\{ t \mid \exists b \in \text{boek} ( \neg t[\text{titel}] = b[\text{titel}] \vee \dots \}$  en bijgevolg is de uitdrukking niet safe.

Naam en voorletters:

Ident.nr.:

3. Druk de volgende vraag over de bibliotheek uit in SQL.  
**Geef de naam en faculteit van alle bibliotheek-gebruikers, met per gebruiker het totaal aantal uitleningen (lopende uitleningen of uitleningen die reeds voorbij zijn).**

Uitwerking:

```
( select u.naam, u.faculteit, count(*)
  from uitlening as u
  group by u.naam, u.faculteit )
union
( select r.naam, r.faculteit, 0
  from reservering as r
  where not exists
    ( select u.naam, u.faculteit
      from uitlening as u
      where u.naam=r.naam and u.faculteit=r.faculteit )
  )
```

**Opmerking:** Je zou het ook met een `outer join` met een “outer join” kunnen oplossen. Doe dit niet! Je moet niet bestuderen en gebruiken wat niet is behandeld, al staat het in het boek.

Probeert deze vraag je om de tuin te leiden? Zoja hoe, en wat heb je gedaan om niet om de tuin geleid te worden?

Als je niet oplet vergeet je de mensen met 0 uitleningen. En als je de bibliotheek-beschrijving niet goed gelezen hebt denk je misschien dat personen worden geïdentificeerd door hun naam, maar dat is naam en faculteit!



Naam en voorletters:

Ident.nr.:

4. Beschouw het volgende stel correcte afleidingsregels voor functionele afhankelijkheden:

**R1:** Reflectiviteitsregel: Als  $Y \subseteq X$  dan  $X \rightarrow Y$ .

**R2:** Verenigingsregel: Als  $X \rightarrow Y$  en  $X \rightarrow Z$ , dan  $X \rightarrow YZ$ .

**R3:** Transitiviteit: Als  $X \rightarrow Y$  en  $Y \rightarrow Z$ , dan  $X \rightarrow Z$ .

Bewijs, dat R1, R2 en R3 (samen) **volledig** zijn.

Bij deze vraag mag je aannemen, dat F1, F2 en F3 (reflexivity, augmentation, transitivity, pag 265 in Silberschatz) correct en volledig zijn (maar je hoeft dit niet te gebruiken als je niet wil).

Om de volledigheid te bewijzen volstaat het om de augmentation rule te bewijzen met behulp van R1, R2 en R3.

Immers, R1 is de reflexivity regel F1 en R3 is de transitivity regel R3.

Stel  $X \rightarrow Y$  geldt en laat  $Z$  een verzameling attributen (uit  $R$ ) zijn.

Uit R1 volgt dat  $XZ \rightarrow Z$  geldt.

Uit R1 volgt ook dat  $XZ \rightarrow X$  geldt.

Uit  $XZ \rightarrow X$  en  $X \rightarrow Y$  volgt met R3 dat  $XZ \rightarrow Y$  geldt.

uit  $XZ \rightarrow Y$  en  $XZ \rightarrow Z$  volgt met R2 dat  $XZ \rightarrow YZ$  geldt.

Waarmee de augmentation bewezen is.

Opgelet, als je het omgekeerde bewijst, namelijk uit F1..F3 R1..R3 afleiden dan bewijs je **soundness** van de regels. Niet wat er is gevraagd, dus fout!

Ook opgelet: je mag een heel bewijs geven vertrekkend van de definitie van functionele afhankelijkheden, maar dat is natuurlijk erg moeilijk.

Nog opgelet: je mag in het bewijs zoals het hierboven staat bij het bewijzen van F2 geen gebruik maken van de deinitie van functionele afhankelijkheden. Je bewijst alles met alleen maar R1, R2, R3, of je bewijst de volledigheid helemaal vanaf de definities, maar geen combinatie van de twee.

Waardering: deelname: 1 punt, vraag 1: 1 punt voor BCNF decompositie, 1 punt voor aanpassing aan nieuwe constraints, vraag 2: 1 punt per query, vraag 3: 2 punten, vraag 4: 2 punten. Dit telt op tot 10.

Geef hier aan wat je zelf denkt dat je cijfer wordt: 10

Optioneel ook deelcijfers (zet cijfer onder de vraag)

1      2a      2b      2c      3      4