Databases 1, Huiswerk 2

Bij dit huiswerk gaan we ervan uit dat je hoofdstuk 3 t/m sectie 3.2 helemaal hebt bestudeerd, en hoofdstuk 7 t/m sectie 7.3.3.

1. Bewijs de correctheid van de "union rule", de "decomposition rule" en de "pseudotransitivity rule", zowel door alleen maar gebruik te maken van de "armstrong axioms" als door alleen maar gebruik te maken van de definitie van functionele afhankelijkheden. (Deze opgave omvat dus exercise 7.8, 7.9 en 7.10 maar nog meer ook.)

   Union Rule met Armstrong regels:
   gegeven: $X \to Y$ en $X \to Z$
   $\Rightarrow$ {Augmentation met $X \to Y$ en X en ook met $X \to Z$ en Y}
   $X \to XY$ en $XY \to YZ$
   $\Rightarrow$ {Transitivity}
   $X \to YZ$

   Union Rule met definitie van fd's:
   Als $X \to Y$ en $X \to Z$ dan geldt in elke toegestane instance dat voor elke t_1 en t_2:
   $t_1[X] = t_2[X] \Rightarrow t_1[Y] = t_2[Y]$ en $t_1[X] = t_2[X] \Rightarrow t_1[Z] = t_2[Z]$
   maar dan ($t_1[Y] = t_2[Y]$ en $t_1[Z] = t_2[Z] \Rightarrow t_1[YZ] = t_2[YZ]$) geldt ook
   $t_1[X] = t_2[X] \Rightarrow t_1[YZ] = t_2[YZ]$.

   Decomposition Rule met Armstrong regels:
   gegeven: $X \to YZ$
   $\Rightarrow$ {Reflexivity op $Y \subseteq YZ$ en op $Z \subseteq YZ$}
   $X \to YZ$ en $YZ \to Y$ en $YZ \to Z$
   $\Rightarrow$ {tweemaal Transitivity}
   $X \to Y$ en $X \to Z$

   Decomposition Rule met definitie van fd's:
   Als $X \to YZ$ dan geldt in elke toegestane instance dat voor elke t_1 en t_2:
   $t_1[X] = t_2[X] \Rightarrow t_1[YZ] = t_2[YZ]$
   maar dan ($t_1[YZ] = t_2[YZ] \Rightarrow t_1[Y] = t_2[Y]$) geldt ook
   $t_1[X] = t_2[X] \Rightarrow t_1[Y] = t_2[Y]$.
   Het tweede deel (voor $X \to Z$) verloopt net zo.

   Pseudo-transitivity Rule met Armstrong regels:
   gegeven $X \to Y$ en $YZ \to T$
   $\Rightarrow$ {Augmentation met $X \to Y$ en Z}
   $XZ \to YZ$ en $YZ \to T$
   $\Rightarrow$ {Transitivity}
   $XZ \to T$

   Pseudo-transitivity Rule met definitie van fd's:
   Als $X \to Y$ dan geldt in elke toegestane instance dat voor elke t_1 en t_2:
   $t_1[X] = t_2[X] \Rightarrow t_1[Y] = t_2[Y]$.
   Alle t_1 en t_2 waarvoor toevallig ook $t_1[Z] = t_2[Z]$ geldt dat zijn ook alle t_1 en t_2 waarvoor $t_1[XZ] = t_2[XZ]$ en voor die tupels geldt $t_1[Y] = t_2[Y]$ en dus ook $t_1[YZ] = t_2[YZ]$.
   De definitie van $YZ \to T$ zegt dat als $t_1[YZ] = t_2[YZ]$ dan geldt $t_1[T] = t_2[T]$.
   Dus geldt voor elke t_1 en t_2 waarvoor $t_1[XZ] = t_2[XZ]$ dat $t_1[YZ] = t_2[YZ]$ en daaruit volgt $t_1[T] = t_2[T]$, wat betekent dat $XZ \to T$.

2. Bewijs dat de volgende regel niet geldt: als in een relatie schema $X \to Y$ en $Z \to Y$ voldaan zijn dan moet $X \to Z$ ook voldaan zijn. (Dit is exercise 7.7.)

   Maak een instance voor R met attributen A, B, C

A	B	C
0	0	0
0	0	1

   Dan geldt $A \to B$ en $C \to B$ maar $A \to C$ geldt niet.

3. Maak opgave 3.5h, met en zonder gebruik te maken van de division operator.
   (Assume the companies may be located in several cities. Find all companies located in every city in which SBC is located.)

   Met division: $\Pi_{\text{company}-\text{name},\text{city}}(\text{company}) \div \Pi_{\text{city}}(\sigma_{\text{company}-\text{name}=\text{"SBC"}}(\text{company}))$

   Zonder division (hiervoor is een standaard formule gegeven op pag. 102, en die laat zien dat er twee keer een verschil nodig is; alle uitwerkingen die niet twee keer een verschil bevatten zijn gegarandeerd fout!):

   $\Pi_{\text{company}-\text{name}}(\text{company})-$
   $\Pi_{\text{company}-\text{name}}($
   $(\Pi_{\text{company}-\text{name}}(\text{company}) \times \Pi_{\text{city}}(\sigma_{\text{company}-\text{name}=\text{"SBC"}}(\text{company})))$
   $-\text{company})$

   Waarom is dit goed? De derde ragel geeft elke company-name samen met elke city waar SBC gevestigd is. In de vierde regel trekken we daar de company-names met de city waar de company zelf gevestigd is van af. We houden dan over elke company-name samen met de cities waar SBC wel gevestigd is en de company niet. Dan projecteren we in de tweede regel op de company-name. Dit zijn precies de foute company-names, dus we trekken ze in de eerste regel van de volledige lijst van company-names af.

4. Stel de volgende vragen over de bibliotheek-database in de relationele algebra (en je mag hierbij de doorsnede, join en toekenning gebruiken als je dat wil):

   (a) Geef de auteurs van wie alle boeken al een keer zijn uitgeleend.
   Let op, hier wordt dus gevraagd dat van elk boek van de auteur tenminste één exemplaar al eens is uitgeleend.
   We lossen dit op via het complement: de auteurs van een boek dat nog niet is uitgeleend.

   $\Pi_{\text{naam},\text{voorletters}}(\text{auteur})-$
   $\Pi_{\text{naam},\text{voorletters}}(\text{auteur} \bowtie (\Pi_{\text{ISBN}}(\text{boek}) - \Pi_{\text{ISBN}}(\text{exemplaar} \bowtie \text{uitlening})))$

   (b) Geef de auteurs van wie alle exemplaren van al hun boeken al een keer zijn uitgeleend.

   $\Pi_{\text{naam},\text{voorletters}}(\text{auteur})-$
   $\Pi_{\text{naam},\text{voorletters}}(\text{auteur} \bowtie \text{exemplaar} \bowtie (\Pi_{\text{barcode}}(\text{exemplaar}) - \Pi_{\text{barcode}}(\text{uitlening})))$
   We zien hier dat we eerst de barcodes zoeken van alle nog nooit uitgeleende exemplaren, dan via een join met exemplaar de ISBN nummers krijgen en dan via een join met auteur de auteurs van die exemplaren. Dit tweede deel geeft de verkeerde auteurs, daarom in de eerste regel nog even een verschil. Wel goed opletten wat er in elke join precies "meegejoind" wordt.

(c) Geef de faculteiten waarvan de leners (samen) van alle faculteiten al wel een keer een exemplaar van een boek hebben geleend.

$\Pi_{\text{uitlening.faculteit,exemplaar.faculteit}}(\sigma_{\text{uitlening.barcode=exemplaar.barcode}}(\text{uitlening} \times \text{exemplaar}))$
$\div \Pi_{\text{exemplaar.faculteit}}(\text{exemplaar})$

De eerste regel zijn de faculteiten van leners met faculteiten van geleende exemplaren (de match gebeurt alleen met barcode!), en we delen door de verzameling van alle faculteiten van exemplaren. We houden dan leners-faculteiten over.
Merk op dat je niet zomaar uitlening met exemplaar mag joinen want dan match je barcode en faculteit.

(d) Geef de faculteiten waarvan een lener al eens van alle faculteiten een keer een exemplaar van een boek heeft geleend.

$\Pi_{\text{uitlening.faculteit}}($
$\Pi_{\text{uitlening.naam,uitlening.faculteit,exemplaar.faculteit}}(\sigma_{\text{uitlening.barcode=exemplaar.barcode}}(\text{uitlening}\times$
$\text{exemplaar})) \div \Pi_{\text{exemplaar.faculteit}}(\text{exemplaar}))$

Dit is bijna hetzelfde, maar de lener moet mee in de deling, en moet daarna worden weggeprojecteerd.

5. Wat beteken de volgende vragen over de bibliotheek-database?
Meta-opmerking: Op deze vraag hebben slechts ongeveer 2% van de studenten een voldoende gehaald (2 van de 3 juist). Het is sterk aan te raden om hierop nog flink te oefenen.

(a) $\Pi_{\text{naam}}(\text{auteur} \bowtie \text{boek} \bowtie \text{reservering})$

Geef de naam van de auteurs waarvan al eens een boek door een naamgenoot is gereserveerd.

(b) $\Pi_{\text{naam}}(\text{auteur} \bowtie \text{boek} \bowtie \text{exemplaar} \bowtie \text{reservering} \bowtie \text{uitlening})$

Geef de naam van de auteurs waarvan al eens een boek door een naamgenoot is gereserveerd en waarvan door die naamgenoot een exemplaar van dat boek is geleend van zijn eigen faculteit.

99% van de studenten hadden deze vraag fout, omdat er meer attributen worden gejoind dan je in de gaten hebt. Met name de faculteit is tricky. Ook hebben velen iets gezegd over "gereserveerd bij zijn eigen faculteit" maar dat kan helemaal niet. Je reserveert een boek, niet een exemplaar. En alleen exemplaren zijn gekoppeld aan een faculteit, boeken niet.

(c) $\Pi_{\text{reservering.naam,reservering.faculteit,exemplaar.faculteit}}$
$(\sigma_{\text{reservering.ISBN=exemplaar.ISBN}}(\text{reservering} \times \text{exemplaar}))$
$\div \Pi_{\text{faculteit}}(\text{exemplaar})$

Geef de naam en faculteit van mensen die wel eens een boek hebben gereserveerd waarvan elke faculteit een exemplaar heeft.

Hier hebben een aantal studenten geschreven "waarvan in elke faculteit een exemplaar aanwezig is" maar dat mag niet want "aanwezig" is een attribuut met een speciale betekenis, en in deze vraag wordt niet geëist dat het exemplaar aanwezig is. Ook bij deze vraag spreken sommigen van "gereserveerd bij zijn eigen faculteit" wat ook hier niet kan.