

MODULE DATABASES

CSG AUGUSTINUS

GEGEVENS BANKEN, SQL & DATASTRUCTUREN

INLEIDING

Ken je iemand wiens bureau eruit ziet zoals in figuur 1? De berg papieren bevat een hoop gegevens of **data**, maar ze leveren de eigenaar van het bureau op dit moment misschien niet de gewenste **informatie**.



FIGUUR 1

De noodzaak om data te structureren wordt groter, naarmate je meer data hebt. Dat geldt in jouw eigen omgeving, maar ook in de digitale wereld. Wie de data niet op een herkenbare manier ordent, heeft geen overzicht. Je kunt er dan geen **informatie** uithalen of nieuwe **kennis** mee vergaren.

Als je de data wel ordent, kun je er **informatie** uithalen: de data krijgt betekenis. Als je gegevens met elkaar weet te verbinden of er patronen in ontdekt, ontwikkel je nieuwe **kennis**. Zo levert één weerstation **data** (in de vorm van de spanning van de weersensoren) over de lokale weersituatie die wordt omgezet naar **informatie** over bijvoorbeeld windkracht, temperatuur en luchtdruk. Meerdere weerstations leveren een totaalbeeld voor een land, dat kan bijdragen aan de **kennis** in de vorm van een weerbericht.

Bij de lessen over programmeren heb je kennis gemaakt met het begrip **variabele**. In een variabele kan een gegeven worden opgeslagen onder een bepaalde naam, zoals *huisnummer* = 4. Een bijzondere variabele was de **array**, waarin je een reeks met gegevens onder één naam kon bewaren. Een voorbeeld hiervan is: *onderwerpVanHetHoofdstuk* = ['html', 'javascript', 'hardware', 'databases'].

Als we dit idee verder uitbreiden komen we uit bij een **tabel**. Op de volgende pagina staan twee tabellen die in het vervolg van dit hoofdstuk vaker zullen worden gebruikt. De tabellen horen bij elkaar. We zeggen dat ze een **relatie** hebben. Een verzameling tabellen die een relatie met elkaar hebben, heet een **relationele database**. Het werken met databases is het hoofdonderwerp van deze module.

Opdracht 1: data, informatie en kennis

1. Wat is het verschil tussen data en informatie? Gebruik de theorie hierboven en internet.
2. In de theorie wordt gesproken over data, informatie en kennis. In deze volgorde neemt de waarde van de gegevens steeds verder toe. Zoek op internet op wat de vierde stap is.
3. Om een voorspelling van het weer van volgende week te kunnen doen, wordt een grote hoeveelheid data, informatie en kennis gecombineerd.
Bedenk twee andere voorbeelden waarbij dit ook het geval is.
4. *Big data* is een actueel item binnen de ICT. Zoek op wat de term betekent.
5. Noem drie bedrijven die gegevens van jou hebben.
6. Bedenk een bedrijf dat iets zou kunnen hebben aan die gegevens. Leg uit waar ze de gegevens voor kunnen gebruiken.
7. Stel dat je een ruim budget hebt om gegevens te kopen. Welke gegevens zou je graag willen hebben? En waarom?

Opdracht 2: informatie uit tabellen halen

Bekijk de tekst op de pagina hiernaast en bestudeer de tabellen. Beantwoord vervolgens de vragen.

8. Van welk type meter bestaan de meeste varianten in de tabel?
9. Welk type meter is het vaakst vervangen? Hoe vaak?
10. Geef het type en het merk van de meter die op dit moment het meest in gebruik is bij de weerstations.
11. Welke meter werd het eerst van allemaal vervangen? En welke werd het snelst vervangen?
Geef het type en het merk en het weerstation waar deze vervanging plaats had.
12. Hoeveel meters die werken op een spanning van 12V worden op dit moment nog gebruikt?
13. Welke beheerders maken op dit moment gebruik van meters die werken op 3V?
14. Bedenk drie vragen die je over deze tabellen zou kunnen stellen. Laat je buurman antwoorden.

Database Weerstations

Hiernaast zie je een tabel met samenwerkende weerstations in de provincie Groningen. Deze tabel **stations** spreekt waarschijnlijk voor zich: hierin staan de verschillende weerstations op een rij met de plaats waar ze zich bevinden en de persoon die het lokale weerstation beheert.

Een weerstation bevat verschillende **meters**. Er is een tabel met gegevens over verschillende typen meters. Hierbij staat ook vermeld van welk merk ze zijn en op welke spanning (voeding) ze werken. Voor elk type meter is er een unieke identificatiecode (*m_id*) gemaakt.

stations		
ws_id	plaats	beheerder
1	Zuidhorn	Van der Veen
2	Groningen	Velthuizen
3	Groningen	Grgurina
4	Bedum	Palsma
5	Garrelsw eer	Osinga

FIGUUR 2

Niet elk weerstation zet dezelfde meters in. Beheerders maken zelf een keuze. Bovendien zijn in de loop der jaren meters kapot gegaan en vervangen door een ander type. Informatie hierover is te vinden in de tabel **inzet**.

Van een bepaald type meter kunnen meerdere exemplaren worden ingezet. Alle exemplaren hebben hun eigen metercode (*m_code*). Dit betekent dat bij één *m_id* meerdere metercodes (*m_code*) kunnen horen. Wanneer een meter is vervangen, is dat te zien doordat in de tabel het veld *eind* is gevuld. In s het veld niet gevuld, dan krijgt deze in de database de waarde **NULL**. De kolom *start* geeft de plaatsingsdatum van een nieuwe meter.

De tabel **inzet** is het lastigst te lezen en te begrijpen. De eerste regel met metercode (*m_code*) 3681 heeft 3 als *ws_id*. Dit is een identificatiecode van het weerstation van beheerder Grgurina uit de tabel **weerstations**.

De meter met de unieke code 3681 heeft metercode (*m_code*) 2. Dit betekent dat het een luchtdrukmeter van het merk Samsung is met een voeding van 12V. Dit blijkt uit de tabel **meters**.

Uit dit voorbeeld blijkt dat deze drie tabellen een relatie met elkaar hebben. Samen vormen de drie tabellen een gegevensbank of **relationele database**.

meters			
m_id	type	merk	voeding
1	luchtdruk	Samsung	5V
2	luchtdruk	Samsung	12V
3	luchtdruk	Philips	5V
4	luchtdruk	Philips	5V
5	luchtdruk	Vavetech	5V
6	luchtdruk	Vavetech	12V
7	luchtdruk	Vavetech	5V
8	windkracht	Vavetech	3V
9	windkracht	Samsung	5V
10	windkracht	Samsung	5V
11	windkracht	Samsung	3V
12	temperatuur	Vavetech	5V
13	temperatuur	Philips	5V
14	temperatuur	Philips	5V
15	temperatuur	Samsung	5V
16	temperatuur	Samsung	12V
17	temperatuur	Samsung	3V
18	temperatuur	Samsung	3V

inzet				
m_code	ws_id	m_id	start	eind
3681	3	2	4-10-2005	28-9-2008
3226	3	8	4-10-2005	
3847	3	13	4-10-2005	7-1-2009
3412	4	2	4-10-2005	9-9-2007
3924	5	2	4-10-2005	30-10-2008
3922	5	12	4-10-2005	
3918	1	5	14-7-2007	
3582	1	7	14-7-2007	
3637	1	12	14-7-2007	
3512	4	4	9-9-2007	20-12-2011
3134	3	6	28-9-2008	
3080	3	12	7-1-2009	
3338	5	15	1-1-2010	1-6-2010
3913	5	8	1-1-2010	
3356	2	9	27-3-2016	
3607	4	5	20-12-2011	
3721	2	1	27-3-2016	
3251	2	15	27-3-2016	
3664	5	7	14-7-2017	

FIGUUR 3

H1 SQL ZOEK EN VIND

1.1 SQL: de computer laten zoeken

Bij opdracht 2 moet je zelf zoeken in de drie tabellen van de database met gegevens van weerstations. Dat wordt al snel een vervelend klusje. Bedenk hoe langdurig jouw zoektocht naar de antwoorden zou zijn geweest, als alle weerstations van Nederland in de tabel waren opgenomen. En als we voor al die weerstations nog meer meetapparatuur zoals regenmeters, luchtvochtigheidsmeters etc. hadden toegevoegd!

Wanneer vraagstukken niet meer te overzien zijn voor mensen of je veel sneller antwoorden wilt krijgen, kun je gebruik maken van de computer. Maar hoe stel je een vraag aan de computer? Hoe leg je de computer uit naar welke gegevens je op zoek bent?

Hiervoor is een speciale taal ontwikkeld: **SQL** of *Structured Query Language*. Met SQL kun je niet alleen een database bevragen om gegevens te selecteren. Je kunt ook gegevens toevoegen, wijzigen of verwijderen en berekeningen uitvoeren, sorteren, groeperen etc. In dit hoofdstuk leer je een aantal van deze technieken om rijen uit een tabel te selecteren. Zo'n rij heet een **record**. Hieronder staan een aantal voorbeelden van sql-zoekopdrachten of **query's**.

SELECT * FROM stations	-- Toont de tabel stations
SELECT merk FROM meters	-- Toont de kolom merk van de tabel meters
SELECT merk,voeding FROM meters	-- Toont zowel de kolom merk als voeding
SELECT beheerder FROM stations WHERE plaats='Groningen'	-- Namen van beheerders in Groningen

In bovenstaande voorbeelden staan typische SQL-uitdrukkingen. Ze zijn hier voor de duidelijkheid met hoofdletters geschreven, maar dat hoeft niet perse. Met **SELECT** geef je aan dat je gegevens wilt selecteren, met **FROM** geef je aan in welke tabel of tabellen de computer moet zoeken. Achter SELECT kun je aangeven welke kolommen (informatie) je terug wilt krijgen. Met de asterisk (*) geef je aan dat je alle kolommen wilt selecteren.

Met **WHERE** kun je een voorwaarde of eis toevoegen. In de laatste query hierboven wordt een lijst met beheerders opgevraagd met de eis dat ze een weerstation in Groningen moeten beheren. Is er meer dan één voorwaarde? Gebruik **AND**: SELECT * FROM meters WHERE type='luchtdruk' AND voeding='5V'

SQL leer je het beste door het te oefenen. Van je leraar krijg je een omgeving met meerdere databases die je kunt gebruiken bij de opdrachten in deze module.

Noteer alle query's en overige antwoorden in een schrift of in een Word-document! Wanneer gevraagd wordt om iets te selecteren, noteer je de query. Wordt er een andere vraag gesteld waarvoor je een query moet gebruiken om het antwoord te vinden, dan noteer je zowel de query als het antwoord.

Opdracht 3: Oefenen met selecteren

Gebruik de database *weerstations*.

15. Voer de query's uit de theorie uit en bekijk het resultaat.
16. Selecteer alle (gegevens van) meters met een 3V-voeding.
17. Welke temperatuurmeter van het merk Samsung werkt op een 5V-voeding? Geef alleen de *m_id*.
18. Gebruik de *m_id* uit de vorige vraag om uit te zoeken hoeveel van deze meters zijn ingezet. Geef de codes (*m_code*), startdatum en einddatum van deze meters.
19. Selecteer alle (gegevens van) luchtdrukmeters van het merk Vavetech met een 5V-voeding.

Opdracht 4: Kennismaken met de database Top-2000

De top-2000 is een jaarlijks terugkerend radioprogramma. Tussen Kerst en Nieuwjaar worden op radio 2 de beste 2000 liedjes aller tijden afgespeeld. Wat die beste liedjes zijn wordt bepaald door mensen die online hun voorkeur aangeven.

Gebruik de database *Top 2000 v1*. Er is ook een database *Top 2000 v2*. Daar gaan we verderop in deze module mee aan de slag. In de database staan de *titel* van het nummer en de naam van de *artiest*. De kolom *uit* is het jaar waarin het nummer is gemaakt. De kolommen *ed2017*, *ed2018*, *ed2019*, *ed2020* en *ed2021* geven de positie van het nummer in de editie van de top 2000 van dat kalenderjaar. De waarde NULL (leeg veld) in één van deze kolommen betekent dat het nummer dat jaar geen notering in de top-2000 had.

20. Voer de query `SELECT * FROM top2000` en bekijk de inhoud van de database.
21. Selecteer het record van het nummer dat op positie 1627 stond in de top-2000 van het jaar 2021?
22. In welk jaar had Maan voor het eerst een notering in de top-2000?
Selecteer alle records met de naam van de artiest om tot je antwoord te komen.
23. Hoeveel nummers van Adele die gemaakt zijn in 2011 hebben een notering gehad in de top-2000?
Selecteer alleen de titel en de kolom *uit*.
24. Welke artiesten hebben een notering gehad in de top-2000 met de titel *Sorry*, maar stonden niet in de editie van 2020? Zoek voor het beantwoorden van de vraag alle liedjes met de titel *Sorry*.
25. Zoek de noteringen van je favoriete artiest en of nummer op.

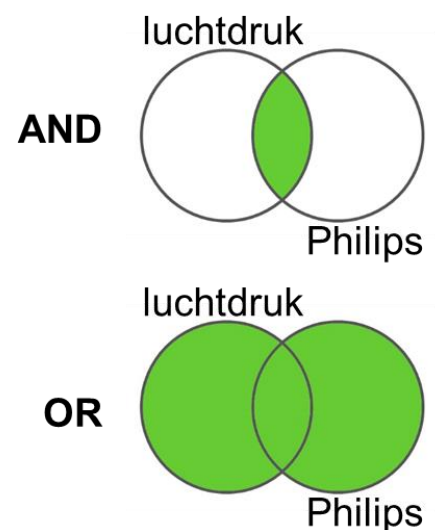
1.2 Logische operatoren

In paragraaf 1.1 hebben we gezien dat we twee voorwaarden kunnen combineren met de uitdrukking **AND**. Dit commando gebruikt je wanneer aan beide voorwaarden tegelijkertijd moet worden voldaan. De Venndiagrammen in figuur 4 tonen het verschil tussen **AND** en **OR**.

De linker cirkel vertegenwoordigt een verzameling meters van het type luchtdruk en de rechter cirkel vertegenwoordigt alle meters van het merk Philip. Het groen gemarkeerde deel van **AND**-tekening bevat dan alle luchtdrukmeters van het merk Philips. Bij **OR** maakt het niet uit: het groen gemarkeerde deel (alles) bevat alle meters die of van het type luchtdruk zijn of van het merk Philips.

Een bijzondere operator is **NOT** wanneer je juist wilt dat aan een eis niet wordt voldaan. En zo zijn er nog veel meer operatoren die je kunt gebruiken. De meeste zullen je bekend voorkomen van de reken- of wiskundelessen en van de programmeerlessen, toen je aan de slag ging met voorwaarden (*if*) en herhalingen (*for* en *while*).

Hieronder zie je voorbeelden van een aantal andere operatoren:



FIGUUR 4

<code>SELECT * FROM meters WHERE voeding='3V' OR voeding='12V'</code>	-- Q1
<code>SELECT * FROM stations WHERE NOT plaats='Groningen'</code>	-- Q2
<code>SELECT * FROM inzet WHERE m_code<3200</code>	-- Q3
<code>SELECT * FROM inzet WHERE NOT eind IS null</code>	-- Q4
<code>SELECT * FROM stations WHERE ws_id>=4</code>	-- Q5
<code>SELECT * FROM meters WHERE NOT (voeding='5V' AND NOT merk='Samsung')</code>	-- Q6 Let op de haakjes!

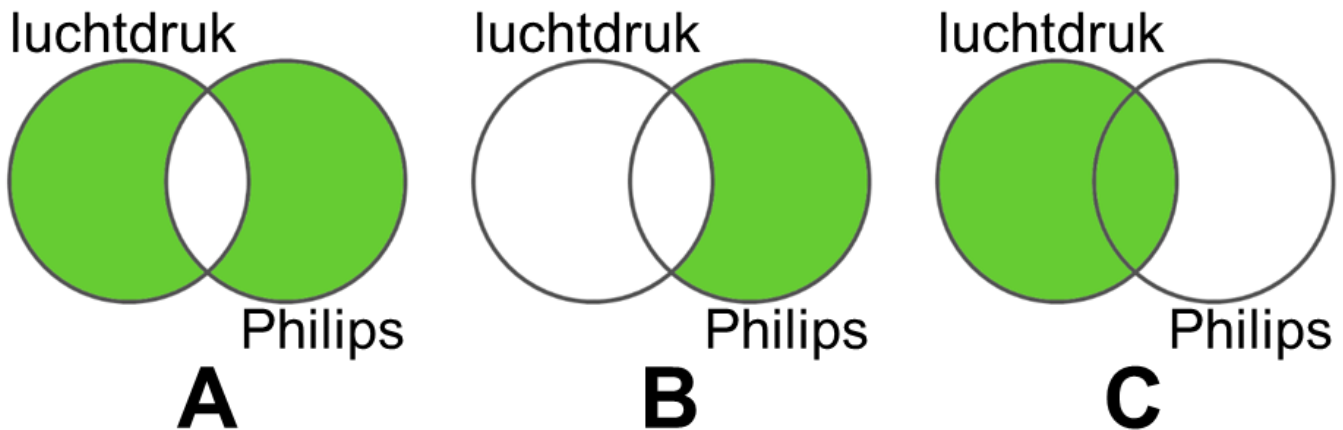
Opdracht 5: Logisch zoeken

Gebruik de database *weerstations*.

26. Lees de query's Q1 t/m Q6 uit de theorie op de vorige bladzijde en voorspel met behulp van de gegeven tabellen van de database *weerstations* wat het resultaat zal zijn.
27. Voer de query's nu uit en vergelijk jouw voorspelling met het resultaat. Begrijp je de query's?
28. Genereer een overzicht van alle windkrachtmeters met een *m_id* die kleiner is dan 10.
29. Genereer een overzicht van alle temperatuurmeters die niet van Samsung zijn.

Extra opdracht 6: nog meer logisch zoeken

Gebruik de database *weerstations*.



FIGUUR 5

30. Bekijk de Venndiagrammen A, B en C in figuur 5. Beschrijf voor A, B en C in woorden welke meters er worden geselecteerd als de Venndiagrammen worden toegepast op de tabel *meters*.
31. Schrijf de query's die horen bij A, B en C.
32. Voer de query's uit. Controleer voor elke query of het resultaat overeen komt met jouw omschrijving en of het resultaat klopt met de gegeven tabel van *meters*.
33. Voorspel wat het resultaat is van de query:
`SELECT * FROM meters WHERE NOT (type='luchtdruk' AND merk='Philips')`
Controleer daarna je antwoord.
34. Voorspel wat het resultaat is van de query:
`SELECT * FROM meters WHERE NOT type='luchtdruk' AND NOT merk='Philips'`
Controleer daarna je antwoord.
35. Voorspel wat het resultaat is van de query:
`SELECT * FROM meters WHERE NOT (type='luchtdruk' OR merk='Philips')`
Controleer daarna je antwoord.

Opdracht 7: Logisch zoeken in de top-2000

Gebruik de database *Top 2000 v1*.

36. Maak een overzicht van de top-2000 van 2017 met achtereenvolgens de kolommen *ed2017*, *artiest* en *titel*. Let goed op: hoeveel records heeft jouw resultaat? (En hoeveel is logisch?)
37. Welk nummer van Oasis stond in de editie van 2017 in de top-100 van de top-2000?
38. Selecteer alle nummers van Anouk die gemaakt zijn na 2009.
39. Selecteer alle nummers van Anouk die tussen 2005 en 2010 uitkwamen (2005 en 2010 doen mee!).
40. Selecteer alle nummers uit de database van Maroon 5 en alle nummers van Stromae.
41. Welke artiesten hebben een liedje gemaakt dat voor 2000 uitkwam en de titel *One* had?
Maak een overzicht met alleen de naam van de artiest en het jaar waarin het liedje is gemaakt.
42. (EXTRA) Hoeveel liedjes uit 1976 stonden in 2011 nog in de top-1500 van de top-2000 maar een jaar later niet meer in de top-1500? Genereer een overzicht van de records. (HINT 4 records is fout)