

DISTINCT in de functie COUNT



Op bladzijde 96 van het boek Databases & SQL wordt een opsomming gegeven van functies die in SQL gebruikt kunnen worden. Bij de COUNT-functie wordt de opmerking gemaakt dat om het aantal verschillende waarden in een kolom te bepalen de frase DISTINCT moet worden opgenomen. Ik zal hieronder aangeven waarom dit zo is en wat het weglaten ervan tot gevolg heeft.

We zullen gebruik maken van de tabel:

<i>Werknemer</i>		
<i>w#</i>	<i>wnaam</i>	<i>salaris</i>
1	Bekx	5000
2	Dijkstra	5000
3	Dijkstra	6000
4	Moens	-
5	Pieters	4500
6	Somers	6500

In de tabel Werknemer is het aantal waarden dat in de kolom wnaam voorkomt 6. Het aantal verschillende waarden is 5 (Dijkstra komt immers 2 maal voor). Willen we het aantal waarden dat voorkomt in de kolom wnaam bepalen dan moeten we *query 1* gebruiken:

```
SELECT COUNT(WNAAM)
FROM WERKNEMER
```

uitkomst:

```
-----
COUNT (WNAAM)
-----
          6
-----
```

Willen we het aantal verschillende namen bepalen dan moeten we *query 2* gebruiken:

```
SELECT COUNT(DISTINCT WNAAM)
FROM WERKNEMER
```

uitkomst:

```
-----
COUNT(DISTINCT WNAAM)
-----
          5
-----
```

Je zou kunnen zeggen waarom gebruik je niet in plaats van de eerste query *query 3*:

```
SELECT COUNT(*)
FROM WERKNEMER
```

uitkomst:

```
-----
COUNT (*)
-----
          6
-----
```

Deze query (query 3) bepaalt echter het aantal rijen van de tabel Werknemer, niet het aantal keer dat een waarde in de kolom wnaam voorkomt. Het verschil hiertussen is te illustreren aan de hand van de kolom salaris. Het aantal rijen in de tabel Werknemer is 6. Het aantal rijen waarin een waarde is opgenomen in de kolom salaris is 5. Het aantal verschillende waarden in de kolom salaris is 4. De queries hiervoor zijn:

- het aantal rijen, zie query 3

- het aantal rijen waar een waarde voorkomt voor salaris (*query 4*):

```
SELECT COUNT(SALARIS)
FROM WERKNEMER
```

uitkomst:

```
-----
COUNT (SALARIS)
-----
          5
-----
```

- het aantal verschillende waarden in de kolom salaris (*query 5*):

```
SELECT COUNT(DISTINCT SALARIS)
FROM WERKNEMER
```

uitkomst:

```
-----
COUNT(DISTINCT SALARIS)
-----
          4
-----
```

Het voordeel van de hier beschreven manier van werken is dat bepaald kan worden in hoeveel rijen een waarde is opgenomen. Dit kan niet uitgaande van de beschrijving van de functie COUNT in het boek.

Een voorbeeld van een omgeving waar de functie COUNT werkt op de hier beschreven manier is SQL bij ORACLE.