

Documentatie SprincFin

Sprincfin helpt bij het vinden van inconsistenties in spreadsheets. Dat is voor een computer een gemakkelijke taak, maar het kan voor mensen lastig zijn om te conceptualiseren wat er nu eigenlijk moet gebeuren. Dit document probeert daarbij te helpen in de vorm van een stappenplan.

1. Bereid de data voor.

In .xlsx-bestanden staat een hoop informatie waar Sprincfin niks mee kan – Sprincfin is alleen geïnteresseerd in de kale data. Deze kale data kan je krijgen door bij Excel te klikken op *Opslaan als*, en dan aan te geven dat je een .txt-bestand wilt. In dit bestand worden dan (1) de data opgeslagen, (2) in welke rij deze data stonden en (3) in welke kolom deze stonden, maar verder niks. Geen cursiveringen, kleuren, etc. Plaats het resulterende .txt-bestand in de map waarin ook Sprincfin draait.

2. Vertel Sprincfin welk bestand hij moet doorzoeken.

Als je Sprincfin opstart, doorzoekt hij zijn eigen map, en geeft hij een lijstje van gevonden .txt-bestanden met daarbij een nummer. Kies het nummer van het bestand dat je wilt gebruiken, en druk op enter.

3. Vertel Sprincfin in welke kolommen inconsistenties kunnen zitten.

Sprincfin weet niet uit zichzelf welke kolommen interessant kunnen zijn en welke niet. Alleen kolommen waar inconsistenties in kunnen zitten zijn interessant, en dat zijn de kolommen waarvan de inhoud samenhangt met andere kolommen. Maak je bijvoorbeeld een database van al je vakantiereizen, dan zouden *vakantieland* en *hoofdstad* interessante kolommen kunnen zijn; *hoofdstad* zou namelijk altijd door *vakantieland* beïnvloed moeten worden. *Kalenderjaar* zou dan weer heel oninteressant zijn, omdat die steeds oploopt en waarschijnlijk geen relatie heeft met een van de andere kolommen (tenzij je, natuurlijk, ieder jaar een nieuw land uitkiest en daar dan meerdere keren heengaat). Je kunt in deze stap aangeven dat maar twee kolommen wilt onderzoeken, maar je kunt er ook heel veel tegelijk doen.

4. Vertel Sprincfin welke kolom ‘bepaalt’ wat een inconsistentie is.

Daarna moet je Sprincfin vertellen vanuit welke kolom hij moet redeneren, welke kolom de categorieën representeert. Stel, je gaat iedere zomer een week naar Italië. Als je dan *vakantieland* en *aantal dagen* bekijkt, en je redeneert vanuit *vakantieland*, dan kan je onder het kopje ‘Italië’ zien of inderdaad alle vakanties naar Italië 7 dagen duurden. Vakanties van andere lengtes staan dan ook onder dit kopje. Bekijk je echter vanuit *Aantal dagen*, dan kan je onder het kopje ‘7’ een overzicht zien van alle vakanties die 7 dagen duurden. Als het goed is staat daar dan heel vaak Italië, maar mogelijk ook andere vakanties.

5. Bekijk en verwerk de output.

Tot slot wil Sprincfin weten hoe hij zijn outputbestand moet noemen. Dit bestand wordt opgeslagen in dezelfde map. Als je op basis van de gegevens in dit outputbestand iets wil aanpassen in het .xlsx-bestand, moet je de desbetreffende rij nog wel terug zien te vinden. Dit kan je gemakkelijk doen door te zoeken (CTRL+F) in het .xls-bestand. Als je bijvoorbeeld in de output ziet dat je een keer per ongeluk 'Pfarijs' in plaats van 'Parijs' hebt getypt, kan je dit makkelijk terugvinden door te zoeken op 'Pfarijs'.