

Inleiding programmeren

1^e jaar wis-, natuur- en sterrenkunde

Universiteit van Amsterdam

oktober 2014

Opgaves bij college 7

data analyse

1 Files inlezen en wegschrijven

Een veel voorkomende toepassing van computerprogramma's is het inlezen en verwerken van grote data bestanden. We zullen vandaag een korte toepassing bekijken hiervan. Hier een paar voorbeelden, maar zoals jullie in de documentatie zien zijn er een aantal manieren om dit te doen.

a) lezen file:

```
input_filehandle = open('inputfile.txt', 'r')
for line in input_filehandle:
    print line
input_filehandle.close()
```

Toegang tot de verschillende parameters in de regel krijg je door de regel in stukken te 'knippen', bijvoorbeeld met behulp van het `split` commando. Het commando `elementen = line.split()` levert een array `elementen` die de losse stukken bevatten. Hierop kan je afzonderlijke bewerkingen uitvoeren.

b) inlezen en uitschrijven file:

```
input_filehandle = open('inputfile.txt', 'r')
output_filehandle = open('outputfile.txt', 'w')
for line in input_filehandle:
    newline = line + " XXXX"
    output_filehandle.write(newline)
input_filehandle.close()
output_filehandle.close()
```

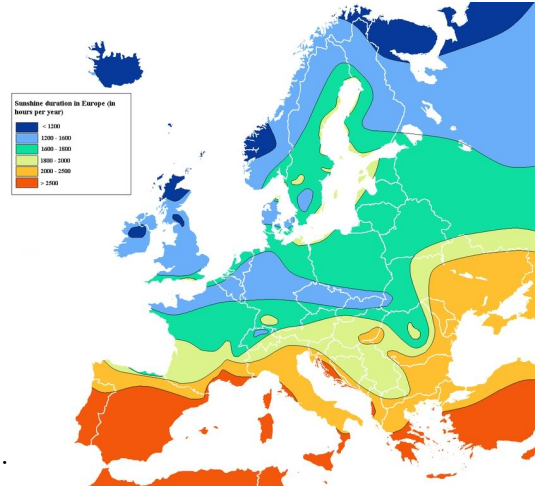
Aanwijzingen voor de opgave van deze week:

In week 5 hebben we gewerkt met de CSV-reader. Dat is in deze week **niet** toegestaan. Het is de bedoeling dat je met de `split()` functie werkt.

opgave 1: klimaatdiscussie

Laten we een steentje bijdragen aan de klimaatdiscussie en data analyseren die door de ECA (European Climate Assessment) beschikbaar wordt gemaakt in grote ASCII files. Data sets zijn hier beschikbaar: ECA data-sets.

We beginnen bescheiden: de temperatuur in De Bilt (station 162). Omdat de data sets groot zijn hebben we die van De Bilt beschikbaar gemaakt op <http://www.nikhef.nl/~ivov/Python/KlimaatData/>.



Download de file `TX_STAID000162.txt` (max. temp.) en `TN_STAID000162.txt` (min. temp.), open ze en lees bovenin hoe de data gecodeerd is. We zien dat de max.(min.) temperatuur op 1 januari 1901 -3.1(-6.8) °C was. Schrijf een programma dat de file doorloopt en beantwoord de volgende vragen.

a) maximumtemperatuur

Wat waren de hoogste en laagste temperatuur die in De Bilt in de 20ste eeuw zijn gemeten ? Op welke dagen was dat ?

b) koud koud koud

Wat is de langste periode dat het aaneengesloten heeft gevroren (maximumtemperatuur onder 0 °C). Wanneer eindigde deze periode ?

hacker edition: Als het 1 dag boven nul geweest mag zijn, wat is dan de langste periode dat het gevroren heeft ? Wanneer eindigde deze ?

opgave 1c: temperatuurverdeling op een specifieke dag

Maak, gebruikmakend van alle jaren van de 20ste eeuw, een grafiek (histogram) van de maximum temperatuur op 13 december, de verjaardag van Martijn Stegeman.