

Inleiding programmeren

1^e jaar wis-, natuur- en sterrenkunde

Universiteit van Amsterdam

oktober 2013

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.

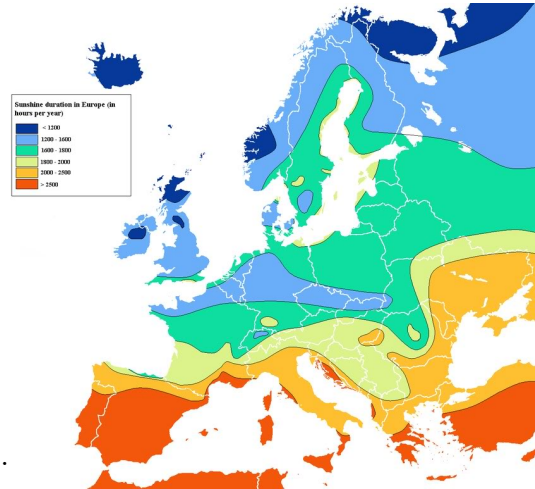
b) inlezen en uitschrijven file:

```
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()
```

Note: In week 5 hebben we gewerkt met CSV. Dat is in deze week **niet** toegestaan. het is de bedoeling dat je met `split()` functie gaat werken.

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 beantwoordt de volgende vragen.

opgave 1a: maximumtemperatuur

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

opgave 1b: elfstedentochten

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: gemiddelde temperatuur per maand

Schrijf een functie die voor een gegeven input (jaar en maand) de gemiddelde max. temp. de output print op het scherm (met de maand in letters i.p.v cijfers en temp. op 2 decimalen nauwkeurig) en als return-value de gemiddelde max. temp. heeft.

opgave 1d: grafieken

Schrijf een functie die, gebruikmakend van de functie uit vraag 1c, een grafiek maakt van de gemiddelde maximum temperatuur per maand in de 20ste eeuw. Maak ook een grafiek (histogram) van de gemiddelde maximumtemperatuur op je verjaardag.

opgave 1e: trends: analyse aantal extreem koude en warme dagen

Bepaal het aantal dagen per jaar dat het extreem koud (min. temp. $< -8^{\circ}\text{C}$) en extreem warm was (max. temp. $> 28^{\circ}\text{C}$) en maak een grafiek van beide als functie van het jaar voor de hele 20^{ste} eeuw. Zie je een trend ?

Hacker: neerslag per seizoen

Op de ECA website kan je heel veel data vinden. Niet alleen temperatuur-data van een zeer groot aantal meetstations net zoals De Bilt, maar ook informatie over het aantal zonne-uren, de hoeveelheid neerslag etc. etc. Een paar weken geleden werd geopperd dat de herfst van 2013 in de top-5 van natste herfsten sinds 1901 zou komen. Deze bewering en de neerslag-gegevens van de top-10 kan je hier vinden: <http://nieuws.weeronline.nl/herfst-2013-top-5-natste-ooit/>.

Onze vraag:

- kloppen de neerslagwaardes van de tabel (top-5) ?
- op welke plek is de herfst van 2013 uiteindelijk terechtgekomen ?