

Python Programming - Labo 5

Doel van dit labo

Deze oefeningen zorgen ervoor dat je van start kan gaan met het ontwikkelen in Python en maken je vertrouwd met de basisbeginselen. In dit labo nemen we de kennis uit de voorgaande labo's mee en introduceren we werken met bestanden, command-line argumenten en foutafhandeling.

OPMERKING: bij bepaalde oefeningen voorzie je zelf een tekstbestand, voorzien van data waarmee je je oefening tot een goed einde kan brengen. In bepaalde oefeningen, 10 tot en met 14, is een dataset voorzien!

Oefening 1

Op Unix gebaseerde besturingssystemen bevatten een tool met de naam head. Het toont de eerste 10 regels van een bestand waarvan de naam is opgegeven als een command-line argument. Schrijf een Python-programma met hetzelfde gedrag. Geef een passend foutbericht weer als het door de gebruiker gevraagde bestand niet bestaat, of als het command-line argument wordt weggelaten.

Oefening 2

Op Unix gebaseerde besturingssystemen bevatten doorgaans ook een tool met de naam tail. Het toont de laatste 10 regels van een bestand waarvan de naam is opgegeven als een command-line argument. Schrijf een Python-programma met hetzelfde gedrag. Geef een passend foutbericht weer als het door de gebruiker gevraagde bestand niet bestaat of als het command-line argument wordt weggelaten. Er zijn verschillende benaderingen die kunnen worden gevolgd om dit probleem op te lossen.

Een optie is om de volledige inhoud van het bestand in een lijst te laden en vervolgens de laatste 10 elementen weer te geven. Een andere optie is om de inhoud van het bestand twee keer te lezen, een keer om de regels te tellen en een tweede keer om de laatste 10 regels weer te geven. Beide oplossingen zijn echter ongewenst bij het werken met grote bestanden. Er bestaat een andere oplossing

waarbij u het bestand slechts één keer hoeft te lezen en slechts 10 regels uit het bestand tegelijk hoeft op te slaan. Voor een extra uitdaging, ontwikkel een dergelijke oplossing.

Oefening 3

Op Unix gebaseerde besturingssystemen bevatten doorgaans ook een tool genaamd `cat`, een afkorting voor concatenate of samenvoegen. Het doel is om de aaneenschakeling van een of meer bestanden weer te geven waarvan de namen worden meegegeven als command-line argumenten. De bestanden worden weergegeven in dezelfde volgorde als waarin ze zijn meegegeven. Maak een Python-programma dat deze taak uitvoert. Je programma genereert ook een passende foutmelding voor elk bestand dat niet kan worden weergegeven en gaat vervolgens verder naar het volgende bestand. Geef een passend foutbericht weer als het programma wordt gestart zonder het command-line argument(en).

Oefening 4

Maak een programma dat regels uit een bestand leest, er regelnummers aan toevoegt en de genummerde regels vervolgens opslaat in een nieuw bestand. De naam van het input-bestand wordt gevraagd aan de gebruiker, evenals de naam van het nieuwe bestand dat je programma zal aanmaken. Elke regel in het output-bestand moet beginnen met het regelnummer, gevolgd door een dubbele punt en een spatie, gevolgd door de regel uit het input-bestand.

Oefening 5

In deze oefening maak je een Python-programma dat de langste woorden in een bestand identificeert. Jouw programma moet een passend bericht weergeven dat de lengte van het langste woord bevat, samen met alle woorden van die lengte die in het bestand voorkwamen. Behandel elke groep karakters die geen witruimte zijn als een woord, zelfs als deze cijfers of leestekens bevat.

Oefening 6

Schrijf een programma dat het woord (of de woorden) weergeeft die het vaakst in een bestand voorkomen. Je programma begint met het vragen van de naam van het bestand dat moet ingelezen worden aan de gebruiker. Dan zou het elke regel in het

bestand moeten verwerken. Elke regel moet worden opgesplitst in woorden en alle leestekens vóór of achter moeten uit elk woord worden verwijderd. Je programma negeert ook de hoofdletters bij het tellen hoe vaak elk woord voorkomt: 'Favoriet' en 'favoriet' bvb zijn dus hetzelfde woord.

Oefening 7

Maak een programma dat alle door de gebruiker ingevoerde getallen bij elkaar optelt. Het negeert daarbij alle invoer die niet uit een geldig getal bestaat (bvb een woord). Je programma geeft de som weer nadat elk nummer is ingevoerd. Het geeft een passend bericht weer telkens de gebruiker iets anders dan een getal heeft ingevoerd. Verlaat het programma wanneer de gebruiker een lege regel invoert. Zorg ervoor dat je programma correct werkt voor zowel integers als floats.

Oefening 8

Python gebruikt een hashtag om het begin van een commentaarregel aan te geven. Deze regel bestaat vervolgens volledig uit commentaar. In deze oefening maak je een programma dat alle opmerkingen uit een Python-bestand verwijdert. Controleer elke regel in het bestand om te bepalen of een #-teken aanwezig is. Als dit het geval is verwijdert het programma alle tekens van het #-teken tot het einde van de regel (we negeren de situatie waarin het commentaarteken ergens in het midden van een regel voorkomt). Sla het gewijzigde bestand op met een nieuwe naam. Zowel de naam van het invoerbestand als de naam van het uitvoerbestand moeten aan de gebruiker worden gevraagd. Zorg ervoor dat er een passend foutbericht wordt weergegeven als zich een probleem voordoet bij het openen van één van de bestanden.

Oefening 9

Schrijf een programma dat een bestand leest dat een lijst met woorden bevat, er willekeurig twee selecteert en ze samenvoegt om een wachtwoord te produceren. Zorg er bij het maken van het wachtwoord voor dat de totale lengte tussen 12 en 15 tekens is en dat elk gebruikt woord ten minste drie letters lang is. Zet elk woord in het wachtwoord met een hoofdletter, zodat de gebruiker gemakkelijk kan zien waar het ene woord eindigt en het volgende begint. Tot slot geeft je programma het wachtwoord weer.

Oefening 10

Schrijf een programma dat een bestand met informatie over chemische elementen inleest en opslaat in een of meer geschikte datastructuren. Vervolgens zou je programma invoer van de gebruiker moeten lezen en verwerken. Als de gebruiker een geheel getal invoert, moet je programma het symbool en de naam van het element met het aantal ingevoerde protonen weergeven. Als de gebruiker een niet-gehele waarde invoert, moet je programma het aantal protonen weergeven voor het element met die naam of dat symbool. Je programma zou een passende foutmelding moeten weergeven als er geen element bestaat voor de naam, het symbool of het aantal ingevoerde protonen. Ga door met het lezen van invoer van de gebruiker totdat er een lege regel is ingevoerd.

Oefening 11

De babynamen-set bevat meer dan 200 bestanden. Elk bestand bevat een lijst met 100 namen, alsook het aantal keer dat de naam voor komt voor elk jaar. De lijnen in de bestanden zijn gerangschikt van meest naar minst vaak gebruikt. Er zijn twee bestanden voor elk jaar: een met namen voor meisjes en de andere met namen voor jongens. De dataset bevat gegevens voor elk jaar van 1900 tot 2012. Schrijf een programma dat elk bestand in de dataset leest en alle namen identificeert die het meest populair waren in ten minste één jaar. Je programma dient twee lijsten weer te geven: een met de meest populaire namen voor jongens en de andere met de meest populaire namen voor meisjes. Geen van de lijsten mag tweemaal dezelfde waarde bevatten.

Oefening 12

Sommige van de namen die voorkomen in de lijsten werden zowel aan jongens als aan meisjes gegeven. Schrijf een programma dat voor een - door de gebruiker ingegeven - specifiek jaar aan meisjes én aan jongens werden gegeven. Je programma geeft een boodschap als je voor dat bepaald jaar geen namen hebt gevonden. Geef een gepaste foutmelding terug als je voor het aangevraagde jaar geen data hebt. Maak gebruik van de babynamen dataset uit oefening 11.

Oefening 13

Gebruik de dataset uit oefeningen 11 en 12. Schrijf een programma waarin de gebruiker twee jaartallen kan ingeven. Analyseer voor deze tijdspanne (bvb van 2002 tot en met 2007) de data en geef de populairste jongens- en meisjesnaam terug voor de volledige periode.

Oefening 14

Gebruik de dataset uit oefeningen 11, 12 en 13. Schrijf een programma dat elk bestand in de dataset uitleest. Terwijl je programma de bestanden uitleest houdt het elke verschillende naam die voorkomt bij (doe dit voor jongens en voor meisjes afzonderlijk). Na afloop rapporteert je programma beide naamlijsten. Deze lijsten bevatten geen dubbele namen.

Oefening 15

Bij het schrijven van een functie is een goede manier van werken om een docstring te voorzien die de werking van de functie beschrijft. Soms vergeet de ontwikkelaar evenwel om de docstring toe te voegen. Maak een programma dat één of meerdere .py bestanden inleest en functies detecteert die niet voorzien zijn van een docstring. Je programma toont alle functies, vergezeld van de bestandsnaam waar de functie zich bevindt. De gebruiker zal de naam of namen van één of meerdere bestanden die moeten nagekeken worden meegeven als command-line argument(en). Je voorziet een gepaste foutmelding wanneer bestanden niet bestaan of kunnen worden geopend. Je programma loopt wel verder en analyseert de bestanden die wel bestaan.