

BI1a-Tinf Thematoets Linux Python I Voorbeeldtoets 2019/2020

Deze toets bestaat uit vier delen opgesplitst in een theoretisch deel en een praktijk deel. Bij het theoretisch deel mag je geen boeken of andere hulpmiddelen gebruiken.

Theoretisch deel

1. Linux theorie (20%, 10 punten)
2. Python theorie (20%, 10 punten)
3. Python Flowchart (10%, 5 punten)

Praktijk deel

4. Python (50%, 25 punten)

Deel 1 Linux

Het Linux en Python analyse deel lever je op papier in. Hierbij is het **niet** toegestaan om boeken of ander naslagwerk te raadplegen. Per volledig goed antwoord krijg je 1 punt.

	Vraag	Commando
1	Commando om de bestanden uit een directory te tonen	
2	Commando om alleen bestanden uit een directory te tonen waar de letter a op positie 2 voorkomt	
3	Commando om de huidige directory te tonen	
4	Commando om de lijst met alle bestanden van een directory op te slaan in bestand met de naam bestanden.txt (het omleiden van output van de terminal naar een bestand)	
5	Het tonen van permissies op bestanden van een directory	
6	Commando om het bestand demo.txt te kopiëren van de huidige directory naar de directory /home/demobestanden/	
7	Het veranderen van de permissies op het bestand demo.txt zodat iedereen recht heeft om het te lezen en te schrijven	
8	Het commando om alle bestanden te verwijderen waar de letter a in voorkomt	
9	Het commando om te laten zien wie er op dat moment ingelogd zijn op de Linux computer	
10	Het commando om het bestand demo.txt te hernoemen tot hello.txt	

Deel 2: Python Theorie

Ieder juist antwoord is 1 punt

```
def main():
    fun1()
    print("4. ", fun2())
    print("5. ", fun2(9))
    print("6. ", fun2(p2=7))
    print("7. ", fun2(1, 9).count("e"))
    print("8. ", fun3(7, 0))
    print("9. ", fun3(7, 7))
    print("10.", fun3("zeven", 10))

def fun1():
    a = 7
    print("1. ", a == 7)
    if a > 5:
        a += 3
        print("2. ", a)
    else:
        print("2. ", a * "a")
    a = 4
    b = 0
    while a < 10:
        a = a + 1
        b = b + 1
    print("3. ", b)

def fun2(p1=7, p2=8):
    if p1 > p2:
        return "groter"
    elif p1 == p2:
        return "gelijk"
    else:
        return "kleiner"

def fun3(p1, p2):
    try:
        getal1 = int(p1)
        getal2 = int(p2)
        uitkomst = int(getal1 / getal2)
        return "abcd"[uitkomst]
    except ZeroDivisionError:
        return "e"
    except ValueError:
        return "f"
    except:
        return "g"

main()
```

<https://pastebin.com/Jdyi0hmG>

Naam: _____

Deel 3: Flowchart

Teken de flowchart voor de functie funforflow(). Volledig juiste flowchart is 5 punten.

```
def funforflow():  
    x = 8  
    if x == 8:  
        if x < 10:  
            print (x)  
        else:  
            print ("hello")  
    else:  
        print ("bye")
```

Deel 4 Python Praktijk

Bij dit deel mag je boeken en digitale bronnen raadplegen. Je mag ook muziek luisteren tijdens het programmeren. Tussendoor kun je rustig koffie gaan drinken en lunchen. Overleg is dan ook prima.

Wat ten strengste verboden is om code met medestudenten uit te wisselen. Dit leidt tot uitsluiting van het tentamen en wordt voorgelegd aan de examencommissie.

Casus

Een onderzoeksinstituut is geïnteresseerd in verschillen tussen genen op verschillende chromosomen. Zoals, wat is het aantal genen op de verschillende chromosomen, gemiddelde GC% en lengte op een chromosoom. Om dit te onderzoeken willen ze een script ontwikkelen dat in staat is deze vragen te beantwoorden.

Het onderzoek richt zich heel breed op allerlei organismen. Als een proof of concept vragen ze je om voor de mens op chromosoom 1 deze functionaliteit te ontwikkelen. Je ontvangt twee bestanden met informatie over de genen bij de mens op chromosoom 1 (chr1.csv en chr1_oefenbestand.csv).

Op de [ensembl website](http://ensembl.org) staat een bestand met alle genen van chromosoom 1 van de mens.

Gene stable ID	Gene name	Gene description	Gene start (bp)	Gene end (bp)	Gene % GC content
ENSG00000223972	DDX11L1	DEAD/H-box helicase 11 like 1 [Source:HGNC Symbol;Acc:HGNC:37102]	11869	14409	57.50
ENSG00000227232	WASH7P	WAS protein family homolog 7, pseudogene [Source:HGNC Symbol;Acc:HGNC:38034]	14404	29570	54.43
ENSG00000278267	MIR6859-1	microRNA 6859-1 [Source:HGNC Symbol;Acc:HGNC:50039]	17369	17436	61.76
ENSG00000243485	MIR1302-2HG	MIR1302-2 host gene [Source:HGNC Symbol;Acc:HGNC:52482]	29554	31109	48.84
ENSG00000284332	MIR1302-2	microRNA 1302-2 [Source:HGNC Symbol;Acc:HGNC:35294]	30366	30503	31.16
ENSG00000237613	FAM138A	family with sequence similarity 138 member A [Source:HGNC Symbol;Acc:HGNC:32334]	34554	36081	50.26
ENSG00000268020	OR4G4P	olfactory receptor family 4 subfamily G member 4 pseudogene [Source:HGNC Symbol;Acc:HGNC:14822]	52473	53312	37.74
ENSG00000240361	OR4G11P	olfactory receptor family 4 subfamily G member 11 pseudogene [Source:HGNC Symbol;Acc:HGNC:31276]	57598	64116	34.27
ENSG00000186092	OR4F5	olfactory receptor family 4 subfamily F member 5 [Source:HGNC Symbol;Acc:HGNC:14825]	65419	71585	32.06
ENSG00000238009	AL627309.1	novel transcript	89295	133723	39.91

Figuur 1: Screenshot van
<http://www.ensembl.org/biomart/martview/42aa5b3d430e2380eff2c404b342bd04>

Van de gegevens uit Figuur 1 is een CSV (Comma Separated Value) bestand gemaakt. In dit bestand zit alle informatie van chromosoom 1. De kolommen zijn gescheiden door comma's. In totaal zijn er 5 kolommen. Aan het eind van iedere regel staat een EOL (End Of Line). Dit is een \n. Dit is een teken, maar is uiteraard niet zichtbaar in de weergave.

Bestand: chr1.csv					
Gene stable ID	Gene name	Gene description	Gene start (bp)	Gene end (bp)	Gene % GC content
ENSG00000223972	DDX11L1	DE...02]	11869	14409	57.50
ENSG00000227232	WASH7P	"WAS pr...GNC:38034]"	14404	29570	54.43
ENSG00000278267	MIR6859-1	microRNA ...HGNC:50039]	17369	17436	61.76

Figuur 2: fragment van het bestand voor analyse. Let op de splitsing van kolommen met een komma en aan het eind een \n als afsluiting van de regel.

Toelichting bestand:

- De eerste regel van het bestand (chr1.csv) bevat de kolomkoppen van de kolommen.
- Er is ook een bestand (chr1_oefenbestand.csv) om je applicatie mee te ontwikkelen. Dit bestand is kleiner en bevat geen kolomkoppen.
- De eerste kolom bevat de unieke identifier voor een gen
- Tweede kolom is een beschrijving van het gen.
- De derde kolom is de start en de vierde het eind van het gen
- De vijfde kolom is het CG% van het gen

In de opdracht beschrijving op staat welke functies je moet schrijven.

Succes!

Opdracht

Schrijf een Python script dat voldoet aan de volgende eisen:

1. Een functie leest het bestand en telt het aantal regels, dit komt overeen met het aantal genen. Functie: **aantal_genen (bestandsnaam)**
Let op: de eerste regel bevat de informatie over de kolommen.
2. Een functie die het aantal genen print met een groter GC% dan het opgegeven GC%. Functie: **groter_gc_perc_dan(bestandsnaam,gc_perc=50)**
3. Een functie die het bestand leest en de gemiddelde lengte van alle genen retourneert (met een return). Functie: **gemiddelde_lengte (bestandsnaam).**
4. Vang **alle** eventuele exceptions af die kunnen optreden. Dus ook exceptions als het lezen buiten een lijst.
5. Voorzie je script van commentaar

Tips

- Werk eerst alle functies uit met het oefenbestand en werk je script daarna zo bij dat het ook werkt met het originele bestand.
- Het originele bestand heeft een titelregel. Deze kun je in je script negeren door een try-except, een while loop of middels een if.
- Alle functies zijn op zichzelf uitvoerbaar.
- Iedere functie opent en leest het bestand weer opnieuw
- Probeer niet alles in een keer op te lossen, maar bekijk eerst het grote geheel en breek het af tot kleine onderdelen
- Probeer iedere keer een klein onderdeel op te lossen
- Als je onverwachte resultaten krijgt print dan tussenresultaten uit
- Het lezen van het laatste element van de lijst (GC%) bevat ook een \n. Deze kun je verwijderen met `rstrip("\n")`

Beoordeling

BI1a-Tinf Beoordelingsformulier Thematoots

		Beoordeling	
		Punten	Student Docent
I.	Code		
	1. Documentatie conform de standaard	5	
II.	Juiste werking:		
	1. Functie aantal_genen (bestandsnaam)	5	
	1. groter_gc_perc_dan(bestandsnaam,gc_perc=50)	5	
	2. gemiddelde_lengte (bestandsnaam)	5	
	3. Exception handling	5	
Eindoordeel		25	

Inleveren

Geschreven code (bestand: **naamstudent_gen.py**) en een eigen beoordeling van je geschreven code aan de hand van bovenstaande beoordelingscriteria. Vul daartoe de kolom student in. N.B. Het eindoordeel van de docent is hetgeen geadministreerd wordt.

Everything in this document was made by the author
 Gonny Henkes-Velemans/Teuntje Peeters/Martijn van der Bruggen
 HAN University of Applied Sciences