EXAMEN: Computergebruik

1e Bachelor Informaticaprof. dr. Peter Dawyndtgroep 1

woensdag 08-01-2014, 8:30 academiejaar 2013-2014 eerste zittijd

Opgave 1 (10 pt)

Gebruik filters, I/O redirection en pipes om telkens een commando samen te stellen dat uitvoer genereert conform onderstaande beschrijvingen. Hierbij is het toegelaten om gebruik te maken van sed, maar niet van andere programmeerbare filters zoals awk, perl, Vermijd dat de commando's (tijdelijke) bestanden aanmaken binnen het bestandssysteem, tenzij dat expliciet gevraagd wordt.

1. Een gegeven tekstbestand (letters.txt in onderstaand voorbeeld) bevat op elke regel een woord, gevolgd door een spatie en een getal $n \in \mathbb{N}_0$. In het woord zelf komen geen spaties voor. Vul de onderstaande commandolijn aan, zodat op elke regel de n-de letter van het woord geselecteerd wordt (de eerste letter heeft index 1) en de gevonden letters achter elkaar worden uitgeschreven op standaard uitvoer.

```
$ cat letters.txt
and 3
now 2
for 2
something 7
completely 8
different 6
$ cat letters.txt | ...
dooier
```

2. Een gegeven tekstbestand (soep1.txt en soep2.txt in onderstaand voorbeeld) bevat een wirwar van afdrukbare karakters die over een aantal verschillende regels staan uitgeschreven. Schrijf een bash shell script lettersoep waaraan de padnaam van een dergelijk tekstbestand als argument moet doorgegeven worden. Het script moet naar standaard uitvoer het woord uitschrijven dat gevormd wordt door alle letters uit het bestand — gelezen van links naar rechts en van boven naar onder — achter elkaar te zetten.

```
$ cat soep1.txt  
,+#62&]|^/&C=;%5[7{/:80_]%o;4\3[,'9_6@>94'\^m>+*?#,39]8_~p<&]9+%!7_,@4$,^=u?~.)0  
/,:?{?+4'%'t+,&;=}[&/'&1"]e-:"%?@r_~/+%?4|{<;:%$$6|~!$\={$^[g'[:e_8+*_342[!&*/','"][;8{b.]%)31:r^2-)!6-.(9}-=|79<92&'8:[.=%76}%ui%!^\-_9'[%0\;+'%k@(3)'&:3*-;~3$  
$ lettersoep soep1.txt  
Computergebruik  
pdawyndt@helios:2013-2014$ cat soep2.txt  
50@[|{@'S^'+<39},:#%!)^6#}.&]='c{1._+3(;$82^!r_[i;!(?))</|;{844?^)'87})7$;;{0'3p$  
:!1.~%{t"=\)3-[}1;${%8'_-+={^'/*-;6'/@0i23=!'%}"~35$!n7_#+-\>).3]!1^%$}!{g<:t8a4}  
'#5>77/"1^@|^~39=2"1+1+#e+'438]=)'$!>!2"}8]_"')n3,[(!/:'.6'4=!^$-\|3>#;/')3<:*;:$ lettersoep soep2.txt  
Scriptingtalen
```

3. Schrijf een bash shell script slingersom waaraan een getal $n \in \mathbb{N}_0$ als argument moet doorgegeven worden. Het script moet naar standaard uitvoer het getal uitschrijven dat bekomen

wordt door de getallen $1, 2, 3, \ldots, n$ achter elkaar te schrijven, afwisselend de operatoren – (optelling) en + (aftrekking) tussen elke twee opeenvolgende cijfers te zetten, en de aldus bekomen uitdrukking uit te rekenen. Voor n=12 krijgen we dan bijvoorbeeld

$$1-2+3-4+5-6+7-8+9-1+0-1+1-1+2=5$$

```
$ slingersom 12
5
$ slingersom 83
14
$ slingersom 182
6
```

4. Schrijf een bash shell script snapshot waaraan als argument de padnaam van een directory moet doorgegeven worden. Het script moet van alle gewone bestanden onder de gegeven directory (inclusief diens subdirectories) die vandaag gewijzigd werden een kopie maken. Deze kopie moet in dezelfde directory geplaatst worden waar ook het originele bestand staat, en moet de oorspronkelijke naam van het bestand overnemen gevolgd door een extensie die de datum van vandaag voorstelt in het formaat JJJJMMDD. Als het bestandssysteem onder de directory aaa bijvoorbeeld de volgende toestand heeft

```
$ ls -lR aaa
aaa:
total 8
drwx----- 2 mpalin users 4096 Jan 6 11:00 bbb
drwx----- 2 mpalin users 4096 Jan 6 11:00 ccc
-rw----- 1 mpalin users 0 Jan 7 12:00 gisteren
-rw----- 1 mpalin users 0 Jan 8 12:00 vandaag1
-rw----- 1 mpalin users 0 Jan 8 12:00 vandaag2
aaa/bbb:
total 0
-rw----- 1 mpalin users 0 Jan 7 12:00 gisteren1
-rw----- 1 mpalin users 0 Jan
                               7 12:00 gisteren2
-rw----- 1 mpalin users 0 Jan
                              8 12:00 vandaag
aaa/ccc:
total 0
-rw----- 1 mpalin users 0 Jan 8 12:00 eergisteren
-rw----- 1 mpalin users 0 Jan 7 12:00 gisteren
-rw----- 1 mpalin users 0 Jan 8 12:00 vandaag
```

dan moet het bestandssysteem onder de directory aaa de volgende toestand hebben na het uitvoeren van het shell script snapshot met als argument de directory aaa.

```
$ snapshot aaa
$ ls -lR aaa
aaa:
total 8
drwx----- 2 mpalin users 4096 Jan 6 11:00 bbb
drwx----- 2 mpalin users 4096 Jan 6 11:00 ccc
                                 7 12:00 gisteren
-rw----- 1 mpalin users 0 Jan
-rw----- 1 mpalin users
                           0 Jan 8 12:00 vandaag1
-rw----- 1 mpalin users
                           0 Jan
                                  8 13:00 vandaag1.20140108
-rw----- 1 mpalin users
                           0 Jan
                                  8 12:00 vandaag2
                           0 Jan 8 13:00 vandaag2.20140108
-rw----- 1 mpalin users
aaa/bbb:
```

```
total 0
-rw----- 1 mpalin users 0 Jan
                               7 12:00 gisteren1
-rw----- 1 mpalin users 0 Jan 7 12:00 gisteren2
-rw----- 1 mpalin users 0 Jan
                              8 12:00 vandaag
-rw----- 1 mpalin users 0 Jan
                               8 13:00 vandaag.20140108
aaa/ccc:
total 0
-rw----- 1 mpalin users 0 Jan 8 12:00 eergisteren
-rw----- 1 mpalin users 0 Jan
                               8 12:00 eergisteren.20140108
                               7 12:00 gisteren
-rw----- 1 mpalin users 0 Jan
-rw----- 1 mpalin users 0 Jan
                               8 12:00 vandaag
-rw----- 1 mpalin users 0 Jan
                               8 12:00 vandaag.20140108
```

Opgave 2 (10 pt)

Het tekstbestand tyfoons.txt bevat informatie over een aantal tyfoons die sinds het jaar 1977 werden waargenomen. De eerste regel van dit bestand bevat de hoofding. Alle andere regels bevatten de volgende informatievelden van één tyfoon: i) regelnummer, ii) identificatienummer, iii) naam, iv) locatie, v) starttijdstip, vi) eindtijdstip, vii) duur en viii) luchtdruk. De informatievelden worden telkens van elkaar gescheiden door één enkele tab. Gevraagd wordt om — gebruik makend van de teksteditors vi of vim — een reeks commando's op te stellen die achtereenvolgens de volgende opdrachten uitvoeren. Probeer voor elke opdracht zo weinig mogelijk commando's te gebruiken en zorg er voor dat elk van deze commando's bestaat uit zo weinig mogelijk tekens. Geen enkele opdracht mag de eerste regel met de hoofding wijzigen, tenzij dit expliciet in de omschrijving van de opgave vermeld staat. Alle opdrachten moeten na elkaar uitgevoerd worden. Ter controle kan je gebruik maken van de meegeleverde bestanden tyfoons.i ($1 \le i \le 5$), die telkens de inhoud van het bestand bevatten nadat de i-de opdracht werd uitgevoerd.

 Vervang het huidige scheidingsteken door een nieuwe reeks scheidingstekens die gevormd wordt door een spatie, gevolgd door een liggend streepje en nog een spatie (-). Plaats ook een hekje (#) voor de eerste regel met de hoofding. Bijvoorbeeld:

```
Basin Birth (UTC)
                                W. N. Pacific
W. S. Pacific
                                                  1977-12-21 00:00
1984-12-19 18:00
                                                                                 1978-01-02 18:00
                                                                                                      12 Days 18 Hours
8 Days 6 Hours
                       MARY
                       FRANK
                                                                                                                       5 Days 18 Hours 945
             201223 SON-TINH
                                         W. N. Pacific 2012-10-23 12:00
4
    344
                                                                                        2012-10-29 06:00
    wordt omgezet naar
    #Rank Number Name Basin Birth (UTC)
                                                           Death (UTC)
                                                                                Duration
                                                                                                   Min Pressure (hPa)
    5 - 197721 - MARY - W. N. Pacific - 1977-12-21 00:00 - 1978-01-02 18:00 - 12 Days 18 Hours - 945 53 - 198402 - FRANK - W. S. Pacific - 1984-12-19 18:00 - 1984-12-28 00:00 - 8 Days 6 Hours - 950
    344 - 201223 - SON-TINH - W. N. Pacific - 2012-10-23 12:00 - 2012-10-29 06:00 - 5 Days 18 Hours - 945
```

2. Pas het veld dat de naam van de tyfoons bevat aan, zodat enkel de eerste letter in hoofdletters staat (alle andere letters moeten desnoods omgezet worden naar kleine letters). De naam moet ook telkens aangevuld worden met een spatie gevolgd door (North) of (South), als het veld met de locatie respectievelijk de tekst W. N. Pacific of W. S. Pacific bevat. Toegepast op het vorige voorbeeld wordt dit

3. Wijzig de volgorde van de velden naar: i) naam, ii) identificatienummer, iii) luchtdruk, iv) duur, v) starttijdstip en vi) eindtijdstip. De velden met het regelnummer en de locatie moeten dus weggelaten worden. Toegepast op het vorige voorbeeld wordt dit

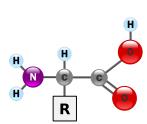
4. De eerste 4 karakters van het veld met het identificatienummer bestaan uit het jaartal, gevolgd door 2 karakters die het rangnummer van de tyfoon binnen dat jaar voorstellen. Pas dit veld aan, zodat er een punt komt te staan tussen jaartal en rangnummer. Toegepast op het vorige voorbeeld wordt dit

```
1 #Rank Number Name Basin Birth (UTC) Death (UTC) Duration Min Pressure (hPa)
2 Mary (North) - 1977.21 - 945 - 12 Days 18 Hours - 1977-12-21 00:00 - 1978-01-02 18:00
3 Frank (South) - 1984.02 - 950 - 8 Days 6 Hours - 1984-12-19 18:00 - 1984-12-28 00:00
4 Son-tinh (North) - 2012.23 - 945 - 5 Days 18 Hours - 2012-10-23 12:00 - 2012-10-29 06:00
```

5. Filter de lijst, zodat naast de hoofdingsregel enkel tyfoons overblijven die een jaarovergang meegemaakt hebben. Je mag ervan uitgaan dat geen enkele tyfoon langer dan 30 dagen duurde. Toegepast op het vorige voorbeeld wordt dit

Opgave 3 (10 pt)

Menselijke eiwitten worden opgebouwd uit 20 verschillende aminozuren. Dit zijn organische verbindingen die allemaal dezelfde generische chemische structuur hebben die hieronder staat weergegeven. De zijketen R is voor elk van de aminozuren verschillend en de polariteit van deze zijketen bepaalt in grote mate de oplosbaarheid van de aminozuren. De gangbare indeling van de aminozuren in zes grote groepen wordt weergegeven in onderstaande tabel. Deze indeling gebeurt op basis van de algemene structuur van de aminozuren en de chemische eigenschappen van hun R-zijketens. Elk aminozuur wordt doorgaans aangeduid met een hoofdletter (hierbij worden alle letters van het alfabet gebruikt, behalve B, J, O, U, X en Z). In de tabel hebben we dan ook zowel de namen als de corresponderende hoofdletters van de aminozuren opgenomen.



KLASSE	NAAM VAN DE AMINOZUREN	LETTERS
alifatisch	glycine, alanine, valine, leucine, isoleucine	$\mathtt{G},\mathtt{A},\mathtt{L},\mathtt{V},\mathtt{I}$
hydroxyl	serine, cysteïne, threonine, methionine	$\mathtt{S},\mathtt{C},\mathtt{T},\mathtt{M}$
cyclisch	proline	P
aromatisch	fenylalanine, tyrosine, tryptofaan	F, Y, W
basisch	histidine, lysine, arginine	H, K, R
zuur	asparaginezuur, glutaminezuur, asparagine, glutamine	D, E, N, Q

Elke regel van het tekstbestand eiwitten.txt bevat de stringvoorstelling van een eiwitsequentie, die bestaat uit een reeks hoofdletters die de aminozuren voorstellen. Daarna volgt telkens één enkele spatie en een woord dat enkel uit hoofdletters bestaat. Gevraagd wordt:

- 1. Bepaal reguliere expressies voor elk van onderstaande verzamelingen. Daarbij staat \mathcal{P} voor de verzameling van alle eiwitsequenties. Probeer de reguliere expressies bovendien zo kort mogelijk te houden.
 - (a) $\alpha = \{ p \in \mathcal{P} \mid p \text{ is een tryptische peptide}^1 \}$
 - ¹ een **tryptische peptide** is een korte eiwitsequentie die eindigt op K of R, en verder geen K of R bevat behalve als die gevolgd wordt door P

voorbeelden: GYWPRPPPNRPMETKPGAR $\in \alpha$, AEAWYYLCTLLVW, VHIGPHEMHKVVNMMVGLR $\not\in \alpha$

(b) $\beta = \{p \in \mathcal{P} \mid \text{in } p \text{ staat na een alifatisch aminozuur nooit een zuur aminozuur (en omgekeerd)} \}$

voorbeelden: ERDYTFQWYCIYQYGLSQ $\in \beta$, KVRAQHYRAMEACVPVGR $\not\in \beta$

(c) $\gamma = \{ p \in \mathcal{P} \mid \text{in } p \text{ staan dezelfde aminozuren steeds minstens twee posities uit elkaar}^2 \}$

² er staan dus minstens twee andere aminozuren tussen

voorbeelden: ANPTQHVQEYPRLDVIMATG $\in \gamma$, YTKPPFHPVAELMPNLEVCM, SHTYPWHWKIFWDTHQGSVF $\not\in \gamma$

(d) $\delta = \{ p \in \mathcal{P} \mid p \text{ bevat een alternerende reeks alifatische en zure aminozuren die minstens 7 aminozuren lang is }$

voorbeelden: TGYFMDKRQLEINIQHEICM, KSQHADGNLQILCEKIAVFY $\in \delta$,

Gebruik een commando uit de grep familie om enkel die regels van het bestand eiwitten.txt te selecteren, waarvan de eiwitsequentie behoort tot de opgegeven verzameling. Vermeld in je antwoordbestand voor elke verzameling het gebruikte selectiecommando, en geef telkens ook aan hoeveel regels je gevonden hebt.

- 2. Beschouw de verzamelingen α , β , γ en δ zoals hierboven gedefinieerd. Gebruik nu deze verzamelingen om op de volgende manier een boodschap bestaande uit vier woorden te achterhalen:
 - (a) het eerste woord staat op de unieke regel met een eiwitsequentie uit de verzameling $\alpha \cap \beta$
 - (b) het tweede woord staat op de unieke regel met een eiwitsequentie uit de verzameling $\beta \cap \gamma$
 - (c) het derde woord staat op de unieke regel met een eiwitsequentie uit de verzameling $\gamma \cap \delta$
 - (d) het vierde woord staat op de unieke regel met een eiwitsequentie uit de verzameling $\delta \cap \alpha$

Vermeld in je antwoordbestand de gevonden woorden, samen met het Unix commando (of de commandosequentie) dat je gebruikt hebt om elk van deze woorden te vinden.

Opgave 4 (10 pt)

Het doel van deze opgave is om een bash shell script te schrijven dat automatisch informatie over een gegeven televisieserie ophaalt uit de Online TV Database (http://thetvdb.com) en deze in een propere opmaak weergeeft in een PDF document. Voor de televisieserie Sherlock moet het PDF bestand dan bijvoorbeeld de volgende informatie weergeven in de aangegeven opmaak.

Sherlock



BBC One 2010-07-25

Sherlock depicts "consulting detective" Holmes, assisting the Metropolitan Police Service, primarily DI Greg Lestrade, in solving various crimes. Holmes is assisted by his flatmate, Dr John Watson, who has returned from military service in Afghanistan. Although the series depicts a variety of crimes and perpetrators, Holmes' conflict with his archnemesis Jim Moriarty is a recurring feature.

Hiervoor ga je als volgt te werk:

1. Maak een LATEX-bestand template.tex dat gebruikt kan worden als template om informatie over een televisieserie weer te geven in tabelvorm. Voor elk informatieveld bevat de template een plaatshouder die achteraf kan vervangen worden door de eigenlijke inhoud van het veld.

Plaatshouders worden aangeduid door een apestaartje (②) gevolgd door de naam van een informatieveld (namen bestaan enkel uit letters). De volgende informatie over een televisieserie zit vervat in de template: i) de naam van de serie (②naam), ii) de locatie van een afbeelding met het logo van de serie (③logo), iii) een korte omschrijving van de serie (④omschrijving), iv) de zender die de serie voor het eerst op televisie heeft uitgezonden (④zender) en v) de datum van die eerste uitzending (②datum). Na compilatie van het LATEX-bestand moet de informatie in een tabel met de volgende opmaak weergegeven worden (voor deze weergave hebben we ②logo reeds vervangen door de locatie van een afbeeldingsbestand). Zorg er voor dat je template de opmaak van de tabel zo getrouw mogelijk weergeeft.



2. Schrijf een bash shell script vervang dat alle tekstregels die binnenkomen via standaard invoer terug uitschrijft naar standaard uitvoer. Hierbij moeten alle voorkomens van @naam in de invoer vervangen worden door het eerste argument dat aan het script wordt doorgegeven, en analoog ook alle voorkomens van respectievelijk @logo, @omschrijving, @zender en @datum door de volgende vier argumenten die aan het script doorgegeven worden. Het script moet dus bijvoorbeeld op de volgende manier kunnen gebruikt worden

```
$ cat template.txt
naam: @naam
logo: @logo
omschrijving: @omschrijving
zender: @zender
datum: @datum
$ vervang aaa bbb ccc ddd eee < template.txt
naam: aaa
logo: bbbb
omschrijving: eee
zender: ccc
datum: dddd</pre>
```

3. Schrijf een bash shell script televisieserie waaraan de naam van een televisieserie als argument moet doorgegeven worden. Het script moet in de eerste plaats de informatie over de opgegeven televisieserie opvragen via de restful API van de Online TV Database. Hiervoor moet het script gebruik maken van de volgende URL template:

```
http://thetvdb.com/api/GetSeries.php?seriesname=naam
```

Als je hierin naam vervangt door de naam van een televisieserie, dan krijg je een XML document terug in het volgende formaat (hier weergegeven voor de serie Homeland).

```
1 | <Data >
2 | <Series >
3 | <seriesid > 247897 </seriesid >
4 | <language > en </language >
5 | <SeriesName > Homeland </SeriesName >
```

```
6
                    <banner>text/247897.jpg</banner>
7
                    <Overview> While in Iraq, Carrie Mathison ...
8
                    <FirstAired>2011-10-02</FirstAired>
9
                    <Network>Showtime</Network>
10
                    <IMDB_ID>tt1796960</IMDB_ID>
11
                    <zap2it_id>EP01466171</zap2it_id>
12
                    <id>247897</id>
            </Series>
13
14
            <Series>
15
                    <seriesid>84450</seriesid>
16
                    <language>en</language>
                    <SeriesName>Homeland Security USA</SeriesName>
17
18
                    <banner>graphical/84450-g.jpg</banner>
19
                    <Overview> Series Gives Viewers ...
20
                    <FirstAired>2009-01-06</FirstAired>
21
                    <Network > ABC </Network >
22
                    <IMDB_ID>tt1245786</IMDB_ID>
23
                    <id>84450</id>
24
            </Series>
   </Data>
```

Omdat er soms meerdere series zijn met dezelfde naam, kan het voorkomen dat het XML document dat je als antwoord terugkrijgt informatie bevat over meerdere televisieseries. Het script moet enkel de informatie van de eerste televisieserie uit de lijst weergeven. Het script moet de naam (SeriesName), de URL van een afbeelding met het logo (banner) en de korte omschrijving (Overview) van deze serie uit het XML document lezen, samen met de zender (Network) en de datum (FirstAired) waarop de serie voor het eerst op televisie werd uitgezonden. De locatie van de afbeelding met het logo van de serie is een relatieve URL met als basis-URL

http://thetvdb.com/banners/

Het script moet daarna een LATEX-bestand aanmaken door het bash shell script vervang uit te voeren om de informatie die uit het XML document geëxtraheerd werd op de juiste plaatsen in te vullen in het LATEX-bestand template.tex. Dit LATEX-bestand moet vervolgens gecompileerd worden tot een PDF document. Enkel dit laatste PDF document mag in de huidige directory opgeslaan worden onder de naam naam.pdf, waabij naam de naam van de televisieserie is die aan het shell script televisieserie werd doorgegeven. Alle andere bestanden die tijdens het uitvoeren van het script aangemaakt worden, moeten opgeslaan worden in de directory tempdir die hiervoor tijdelijk wordt aangemaakt in de huidige directory. Na het uitvoeren van het shell script moet deze tijdelijke directory met zijn volledige inhoud terug verwijderd zijn. Zorg er ook voor dat het shell script geen enkele uitvoer uitschrijft naar het standaard uitvoerkanaal en naar het standaard errorkanaal.

Belangrijk: Genereer de PDF bestanden sherlock.pdf en prison break.pdf door het shell script televisieserie als volgt aan te roepen. Plaats deze PDF bestanden in het ZIP-bestand dat je indient via Indianio.

```
$ televisieserie sherlock
$ televisieserie "prison break"
```