
EXAMEN: Computergebruik

1^e Bachelor Informatica
prof. dr. Peter Dawyndt
groep 1

woensdag 24-08-2011, 8:30
academiejaar 2010-2011
tweede zittijd

Opgave 1

Het tekstbestand `gemeentes.txt` bevat informatie over alle gemeentes in het Vlaams gewest. Elke regel bevat de volgende gegevens over één bepaalde gemeente: *i*) naam, *ii*) stad of gemeente, *iii*) provincie, *iv*) oppervlakte in km², *v*) aantal inwoners en *vi*) lijst van deelgemeentes. De verschillende informatievelen worden van elkaar gescheiden door een dubbelpunt (:). De hoofding van het tekstbestand bestaat daarnaast uit een aantal commentaarregels die beginnen met een hekje (#). Gevraagd wordt om, gebruik makend van de teksteditor `vi` of `vim`, een reeks commando's op te stellen die achtereenvolgens de volgende opdrachten uitvoeren:

1. Zorg er voor dat alle velden gescheiden worden door een puntkomma (;), en dat alle spaties vooraan en achteraan elk veld verwijderd worden. Zorg er ook voor dat de tekst 'gem.' vervangen wordt door de tekst 'gemeente'. Zo moet bijvoorbeeld

```
Beringen:stad:Limburg:78,30:42758:Beringen:Beverlo: Koersel:Paal  
Brasschaat : gem.:Antwerpen:38,49:36949:Brasschaat
```

omgezet worden naar

```
Beringen;stad;Limburg;78,30;42758;Beringen;Beverlo;Koersel;Paal  
Brasschaat;gemeente;Antwerpen;38,49;36949;Brasschaat
```

2. Zorg er voor dat de deelgemeentes worden opgelijst met een komma en een spatie als scheiding, in plaats van als verschillende velden. Toegepast op het vorige voorbeeld wordt dit:

```
Beringen;stad;Limburg;78,30;42758;Beringen, Beverlo, Koersel, Paal  
Brasschaat;gemeente;Antwerpen;38,49;36949;Brasschaat
```

3. Vervang bij gemeentes met een oppervlakte groter of gelijk aan 92 km² de aanduiding 'gemeente' door 'grote gemeente' en bij steden met een oppervlakte kleiner of gelijk aan 38 km² de aanduiding 'stad' door 'kleine stad'. Een aantal voorbeelden van het resultaat:

```
Mol;grote gemeente;Antwerpen;114,26;34114;Mol  
Mesen;kleine stad;West-Vlaanderen;3,58;950;Mesen  
Beringen;stad;Limburg;78,30;42758;Beringen, Beverlo, Koersel, Paal  
Brasschaat;gemeente;Antwerpen;38,49;36949;Brasschaat
```

4. Wissel het veld gemeente/stad van plaats met het veld provincie. Laat het veld met de oppervlakte van de gemeente weg. Toegepast op het vorige voorbeeld wordt dit:

```
Mol;Antwerpen;grote gemeente;34114;Mol  
Mesen;West-Vlaanderen;kleine stad;950;Mesen  
Beringen;Limburg;stad;42758;Beringen, Beverlo, Koersel, Paal  
Brasschaat;Antwerpen;gemeente;36949;Brasschaat
```

5. Voeg na de naam van de gemeente een extra veld toe dat de postcode van de gemeente bevat. Gebruik hiervoor het `join` commando en het bestand `postcodes.txt`. Zorg er ook voor dat de gemeentes gesorteerd worden op naam. Toegepast op het vorige voorbeeld wordt dit:

Beringen;3580;Limburg;stad;42758;Beringen, Beverlo, Koersel, Paal
Brasschaat;2930;Antwerpen;gemeente;36949;Brasschaat
Mesen;8957;West-Vlaanderen;kleine stad;950;Mesen
Mol;2400;Antwerpen;grote gemeente;34114;Mol

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. De commentaarregels van de hoofding mogen door je commando's niet gewijzigd worden, zelfs niet als er bijkomende regels aan de hoofding worden toegevoegd. Alle wijzigingen moeten na elkaar uitgevoerd worden.

Opgave 2

Een *rainbow-tabel* bevat een lijst van mogelijke wachtwoorden en de berekende hashes van deze wachtwoorden. Dergelijke tabellen worden bijvoorbeeld gebruikt om wachtwoorden te testen op hun veiligheid, of om ze te kraken. Het bestand `rainbow.txt` is een voorbeeld van een rainbow-tabel. Elke regel bevat de md5-checksum van een mogelijk wachtwoord, gevolgd door een spatie en het wachtwoord zelf. Wachtwoorden bevatten nooit spaties, en een md5-checksum bestaat enkel uit cijfers en de hoofdletters A tot en met F. Gevraagd wordt:

1. Bepaal reguliere expressies voor elk van onderstaande verzamelingen, waarbij \mathcal{H} de verzameling van alle md5-checksums voorstelt. Probeer deze reguliere expressies zo kort mogelijk te houden.
 - (a) $\alpha = \{h \in \mathcal{H} \mid \text{derde karakter van } h \text{ is gelijk aan derde laatste karakter van } h\}$
 - (b) $\beta = \{h \in \mathcal{H} \mid h \text{ bevat vier opeenvolgende cijfers, die bestaan uit twee paren gelijke cijfers}\}$
voorbeelden: 1221, 3344, 5665, 7777
 - (c) $\gamma = \{h \in \mathcal{H} \mid \text{elke letter komt maximaal tweemaal voor in } h\}$
 - (d) $\delta = \{h \in \mathcal{H} \mid \text{aantal cijfers in } h \text{ is een veelvoud van vier}\}$

Gebruik een commando uit de `grep` familie om enkel die regels uit `rainbow.txt` te selecteren die behoren 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 deze verzamelingen om op de volgende manier een geheime boodschap bestaande uit vier woorden te achterhalen die in de rainbow-tabel verborgen zit:
 - (a) het eerste woord is het wachtwoord op de unieke regel waarvan de md5-checksum $\in \alpha \cap \beta$
 - (b) het tweede woord is het wachtwoord op de unieke regel waarvan de md5-checksum $\in \beta \cap \gamma$
 - (c) het derde woord is het wachtwoord op de unieke regel waarvan de md5-checksum $\in \gamma \cap \delta$
 - (d) het vierde woord is het wachtwoord op de unieke regel waarvan de md5-checksum $\in \delta \cap \alpha$

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. Elk commando of elke commandosequentie moet dus als resultaat één van de gezochte woorden naar standaard uitvoer schrijven (zonder de md5-checksum die aan het woord voorafgaat).

Opgave 3

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

1. Het `history` commando toont een lijst van de laatst gebruikte commando's. Gebruik dit commando om een nieuw commando op te stellen, dat je tien meest favoriete commando's naar standaard uitvoer schrijft. Opties en argumenten van de gebruikte commando's in de `history` lijst moeten hierbij genegeerd worden, en bij samengestelde commando's (pipes) moet je enkel het eerste commando weerhouden bij het uitvoeren van de telling. Zorg er voor dat de commando's in het overzicht worden weergegeven volgens dalend aantal keer gebruikt, en in tweede instantie alfabetisch gerangschikt op naam. Onderstaande sessie illustreert hoe de uitvoer van je commando er moet uitzien:

```
$ history | ...
145 cat
 38 pico
 33 echo
 33 ls
 30 mv
 26 find
 26 gawk
 15 history
 14 python
 12 cd
```

2. Op de HTML-pagina met URL <http://www.python.org/download/> kan je de standaardimplementatie van Python downloaden. De broncode van de huidige productiesversies (er kunnen er meerdere zijn) kan je hier bijvoorbeeld downloaden als een tarball die gecomprimeerd werd met het `bzip` commando. Dergelijke bestanden zijn te herkennen aan de extensie `.tar.bz2`. De bestanden die je hier kunt downloaden, gebruiken als sjabloon voor de naam `Python-x.y.z.tar.bz2`, waarbij `x.y.z` het versienummer van de huidige productversie aangeeft. Gebruik deze informatie om een commando te schrijven dat de webpagina ophaalt, daarin de namen van de gecomprimeerde tarballs opzoekt, en er het versienummer uitfiltert. Het commando moet op die manier de versie nummers van de huidige productversies van Python naar standaard uitvoer schrijven:

```
$ commando
2.7.2
3.2.1
```

3. Het bestand `multi.fasta` maakt gebruik van het zogenaamde *FASTA formaat*. Dit is een tekstgebaseerd bestandsformaat dat gebruikt wordt in de bioinformatica om DNA of eiwitsequenties op te slaan. Individuele baseparen of eiwitresidu's worden daarbij voorgesteld door één-letter codes. Het formaat laat ook toe om de verschillende sequenties te laten voorafgaan door sequentiename en andere informatie velden.

Een sequentie in FASTA formaat begint met een één-regel beschrijving, gevolgd door de eigenlijke sequentiegegevens die eventueel kunnen gesplitst worden over verschillende regels. De regel met de beschrijving wordt onderscheiden van de sequentiegegevens door een “groter dan” symbool (`>`) in de eerste kolom. Het woord dat volgt op het `>` symbool is de identifier van de sequentie, en de rest van de regel is de beschrijving (beide zijn optioneel). Er mag geen spatie staan tussen het `>` symbool en de eerste letter van de identifier en het is aanbevolen dat alle regels korter zijn dan 80 lettertekens. Elke sequentie eindigt waar een nieuwe regel begint met een `>` symbool,

wat de start van een nieuwe sequentie aangeeft, of op het einde van het bestand. Een eenvoudig voorbeeld van één enkele sequentie in FASTA formaat:

```
>118480563|DQ207729|Bacillus cereus|16S ribosomal RNA gene, complete sequence
AGAGTTTGATCTTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAATGGATTA
AGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAACACGTTGGTAACTGCCATAAGACTGGG
ATAACTCCGGGAAACCGGGGCTAATACCGGATAACATTTGAACCGCATGGTTCGAAATTGAAAGGCGGC
TTCGGCTGTCACTTATGGATGGACCCGCTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAA
CGATGCGTA
```

De beschrijvingsregel — de regel die begint met een > symbool — geeft een naam en/of een unieke identifier aan de sequentie, en bevat vaak ook nog andere informatie velden. Informatie velden worden dan van elkaar gescheiden door een verticale streep (|). Verschillende sequentiedatabanken maken gebruik van gestandaardiseerde beschrijvingsregels, wat helpt om automatisch de informatie uit de regels te kunnen extraheren. Een voorbeeld van een FASTA bestand met meerdere sequenties, gegenereerd voor de GenBank databank van het NCBI, ziet er als volgt uit:

```
>571435|U16165|Clostridium acetobutylicum|NCIMB8052 16S ribosomal RNA (rrn) gene
TGGCGGCGTGTAAACACATGCAAGTCGAGCGATGAAGCTCCTTCGGGAGTGGATTAGCGGCGGACGGGT
GAGTAAACACGTGGGTAACCTGCCTATAGAGGGGAATAGCCTTTTCAAAGGAAGATTAATACCGCATAAG
ATTGTAGTGCCGCATGGCATTAAAGGAGTAATCCGCTATGAGATGGACCCGCTCGCATTAGCT
AGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACATTGG
GACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTG
>996091|L07834|Geobacter metallireducens|16S ribosomal RNA gene
AGAGTTTGATCCTGGCTCAGAACGAACGCTGGCGGAGTGCCTAACACATGCAAGTCGAACGTGAAGGGGG
CTTCGGTCCCGGAAAGTGGCGCACGGGTGAGTAACCGGTGGATAATCTGCCAGTGATCTGGGATAACA
TCTCGAAAGGGGTGCTAATACCGGATAAGCCACGGAGTCCTTGGATTCTCGGGAAAAAGGGGGGACCT
TCGGGCTTTTGTCACTGGATGAGTCCGCGTACCATTAGCTAGTTGGTGGGGTAATGGCCACCAAGGCT
ACGATGGTTAG
```

We gaan er bij aanvang van uit dat de huidige directory het bestand `multi.fasta` bevat. Gevraagd wordt:

- (a) Geef één enkel commando dat een nieuwe directory `fasta` aanmaakt onder de huidige directory, en er tevens voor zorgt dat de directory `fasta` de nieuwe huidige directory wordt.
- (b) Geef een commando dat het bestand `multi.fasta` — dat verschillende sequenties bevat — opsplijst in verschillende bestanden die elk één enkele sequentie in FASTA formaat bevatten. Deze bestanden moeten in de huidige directory geplaatst worden, en moeten benoemd worden als `fastan`, waarbij `n` een volgnummer is. Zorg er voor dat er geen informatie naar de terminal geschreven wordt bij het uitvoeren van het commando. Onderstaande sessie illustreert de inhoud van de huidige directory nadat het commando werd uitgevoerd:

```
$ ls
fasta00  fasta01  fasta02  fasta03
$ cat fasta00
>118480563|DQ207729|Bacillus cereus|16S ribosomal RNA gene, complete sequence
AGAGTTTGATCTTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAATGGATTA
AGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAACACGTTGGTAACTGCCATAAGACTGGG
ATAACTCCGGGAAACCGGGGCTAATACCGGATAACATTTGAACCGCATGGTTCGAAATTGAAAGGCGGC
TTCGGCTGTCACTTATGGATGGACCCGCTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAA
CGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGA
GGCAGCAGTAGGGAATCTTCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCTGAGTGATGAAGGCT
```

- (c) Schrijf een `bash` shell script `hernoem.sh` dat eveneens in de huidige directory geplaatst wordt. Aan dit shell script moet één enkel bestand als argument doorgegeven worden dat één enkele sequentie in FASTA formaat bevat. Het shell script moet het bestand hernoemen op basis van het zogenaamde *accession number* dat als tweede veld op de informatieregel van het FASTA bestand staat, aangevuld met de extensie `.fasta`.

```

$ ls
fasta00    fasta01    fasta02    fasta03
$ hernoem.sh fasta00
$ ls
DQ207729.fasta    fasta01    fasta02    fasta03

```

- (d) Geef een commando dat het `bash` shell script `hernoem.sh` toepast op alle bestanden met prefix `fasta` in de huidige directory.

```

$ ls
DQ207729.fasta    fasta01    fasta02    fasta03
$ commando
$ ls
DQ207729.fasta    L07834.fasta    U16165.fasta    U86373.fasta

```

Opgave 4

1. Geef \LaTeX -code die een reconstructie maakt van onderstaande tabel, samen met de bijhorende tekst. Zorg er daarbij voor dat de opmaak zo getrouw mogelijk behouden blijft.

De tabel

Product	1	2	3	4	5
Prijs	€ 124	€ 136	€ 85	€ 156	€ 23
Garantie {jaar}	1	2	-	3	1
Score	89%	84%	51%		45%
Aanbevolen	ja	ja	neen	neen	neen

bevat een vergelijking van de verschillende types RAM geheugen.

2. Geef \LaTeX -code die hetzelfde resultaat oplevert als het tekstfragment in onderstaand kader.

De *Hessiaan* van een functie in meerdere veranderlijken $f(x_1, x_2, \dots, x_n)$ is de matrix van tweede-orde partiële afgeleiden van die functie, en is dus gedefinieerd als:

$$\nabla^2(x_1, x_2, \dots, x_n) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix} = (\text{kortweg}) \left(\frac{\partial^2 f}{\partial x_1 \cdot \partial x_2} \right).$$

De benaming *Hessiaan* verwijst naar de Duitse wiskundige Ludwig Otto Hesse en is naar verluidt geïntroduceerd door de Engelse wiskundige James Joseph Sylvester.

3. Geef \LaTeX -code die hetzelfde resultaat oplevert als het tekstfragment in onderstaand kader.

De gemiddelde intensiteit $J_\lambda(z)$ wordt gedefinieerd als het gemiddelde van de specifieke intensiteit $I_\lambda(z, \mu)$ over alle ruimtehoeken,

$$\begin{aligned} J_\lambda(z) &= \oint I_\lambda(z, \mu) \frac{d\Omega}{4\pi} \\ &= \frac{1}{4\pi} \int_0^{2\pi} d\varphi \int_0^\pi I_\lambda(z, \mu) \sin \theta d\theta, \end{aligned} \quad (1)$$

hetgeen zich herleidt tot

$$J_\lambda(z) = \frac{1}{2} \int_{-1}^1 I_\lambda(z, \mu) d\mu \quad (2)$$

aangezien de specifieke intensiteit axiaal symmetrisch verondersteld werd.

Plaats een PDF bestand met daarin de gecompileerde L^AT_EX fragmenten in het ZIP-bestand dat je indient via Indianio.