
EXAMEN: Computergebruik

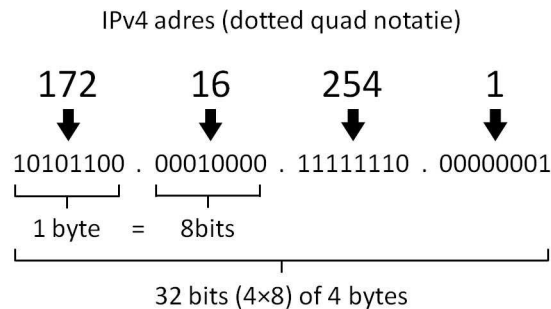
Prof. Dr. Peter Dawyndt
1^e Bachelor Informatica
groep 1

maandag 17-08-2009, 14:00h
academiejaar 2008-2009
tweede zittijd

Opgave 1

In computernetwerken wordt vaak gebruik gemaakt van de zogenaamde *dotted quad* notatie (ook gekend als *quad-dotted* notatie of *dot-decimal* notatie). Dit is een method om binaire getallen te schrijven in per octet gegroepeerde decimale getallen, waarbij de decimale getallen die corresponderen met elke groep van elkaar worden gescheiden door een punt.

Zoals aangegeven in onderstaande figuur worden IPv4 adressen bijna uitsluitend voorgesteld in dotted quad notatie. Een IPv4 adres bestaat dan uit vier natuurlijke getallen kleiner dan 256 die van elkaar worden gescheiden door drie punten. Zo wordt bijvoorbeeld het hexadecimale getal `0xFF000000` in dotted quad notatie geschreven als `255.0.0.0`.



Het bestand `IPv4.txt` bevat een lijst van IPv4 adressen in dotted quad notatie. Elk IPv4 adres staat op een afzonderlijke regel, gevolgd door een spatie en een woord. Gevraagd wordt:

1. Bepaal reguliere expressies voor elk van de onderstaande verzamelingen, waarbij \mathcal{I} de verzameling van alle geldige IPv4 adressen in dotted quad notatie voorstelt. Probeer deze reguliere expressies zo kort mogelijk te houden.
 - (a) $\alpha = \{i \in \mathcal{I} \mid \text{geen enkel natuurlijk getal in } i \text{ bestaat enkel uit twee cijfers}\}$
 - (b) $\beta = \{i \in \mathcal{I} \mid \text{eerste en derde natuurlijk getal in } i \text{ zijn gelijk}\}$
 - (c) $\gamma = \{i \in \mathcal{I} \mid \text{som van aantal vieren en aantal vijven in } i \text{ is even en groter dan nul}\}$
 - (d) $\delta = \{i \in \mathcal{I} \mid i \text{ bevat minstens 3 zessen maar } i \text{ bevat geen opeenvolgende zessen}\}$

opmerking: zessen gescheiden door een punt worden ook als opeenvolgend beschouwd

Gebruik een commando uit de `grep` familie om enkel die regels van het bestand `IPv4.txt` te selecteren die beginnen met een IPv4 adres in dotted quad notatie dat tot de opgegeven verzameling behoort. Vermeld in je antwoordbestand voor elke verzameling het gebruikte selectiecommando, en geef 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 IPv4 adres uit verzameling $\alpha \cap \beta$.

- (b) Het tweede woord staat op de unieke regel met IPv4 adres uit verzameling $\beta \cap \gamma$.
- (c) Het derde woord staat op de unieke regel met IPv4 adres uit verzameling $\gamma \cap \delta$.
- (d) Het vierde woord staat op de unieke regel met IPv4 adres uit verzameling $\delta \cap \alpha$.

Vermeld in je antwoordbestand de gevonden woorden, samen met de regelnummers in het bestand `IPv4.txt` waarop deze woorden gevonden werden. Geef voor elk woord ook het `grep` commando (of de commandosequentie) die je gebruikt hebt om het woord en het regelnummer te vinden.

Opgave 2

Het bestand `cyclonen.txt` bevat een lijst van de natste tropische stormen per land, verzameld in Wikipedia op basis van een aantal verschillende bronnen. Elke regel bevat gegevens over één enkele storm zoals aangegeven in onderstaand formaat, waarbij de velden worden gescheiden door één enkel dubbelpunt (:).

```
<land>:<regenval (mm)>:<regenval (inches)>:<naam>:<meetstation>
```

Het bestand bevat bovendien bovenaan nog enkele commentaarregels die beginnen met een hekje (#). Omdat deze gegevens rechtstreeks uit Wikipedia komen, willen we het gebruikte bestandsformaat nog enigszins aanpassen. Daarom wordt gevraagd om, gebruik makend van de teksteditor `vi` (of `vim`), een reeks substitutiecommando's op te stellen die achtereenvolgens de volgende opdrachten uitvoeren:

1. Zorg ervoor dat alle spaties vooraan en achteraan elk veld verwijderd worden.
2. Sommige velden bevatten overtollige informatie die ook moet verwijderd worden. Verwijder de tekst " mm" uit het derde veld en de tekst " inches" uit het vierde veld indien die daar mocht voorkomen. Bovendien bevatten de velden mogelijks ook literatuurverwijzingen onder de vorm van een natuurlijk getal tussen vierkante haakjes. Deze literatuurverwijzing moeten ook verwijderd worden. Zo wordt bijvoorbeeld

```
Mozambique:1:502 mm:19.76 inches:Eline 2000:Levubu[163]
```

omgezet naar

```
Mozambique:1:502:19.76:Eline 2000:Levubu
```

3. Het eerste veld bevat voor de naam van het land soms ook nog de naam van een plaats binnen dat land (meestal een eiland), waarbij beide namen van elkaar worden gescheiden door een komma gevolgd door een spatie. Zorg ervoor dat in het eerste veld enkel de naam van het land wordt behouden. Zo moet bijvoorbeeld `Sint Maarten, Netherland Antilles` worden omgezet naar `Netherland Antilles`.
4. Reorganiseer het bestand zodat de regels geordend zijn op land (alfabetisch) en daarna op hoeveelheid neerslag (in mm) die de storm met zich mee heeft gebracht (waarbij stormen met de hoogste regenval bovenaan komen te staan). Zorg ervoor dat de regels na herorganisatie dezelfde vorm behouden. Aangezien dit het veld met het rangnummer overbodig maakt, moet ook het tweede veld integraal verwijderd worden.
5. De naam van de storm bevat meestal een jaartal dat bestaat uit vier cijfers. Extraheer het jaartal uit de naam van de storm en voeg het als een nieuw veld toe achter de naam van storm. Zorg ervoor dat aan regels die geen jaartal in de naam van de storm bevatten ook een nieuw leeg veld wordt toegevoegd op de corresponderende plaats. Zo wordt bijvoorbeeld

```
Australia:927:36.50:April 1898 cyclone:Whim Creek
```

```
Belize:249.2:9.81:Chantal 01: Towerhill
```

omgezet naar

```
Australia:927:36.50:April 1898 cyclone:1898:Whim Creek
```

```
Belize:249.2:9.81:Chantal 01::Towerhill
```

6. Markeer alle stormen die 1000 mm of meer neerslag produceerden door op het einde van de corresponderende regel de tekst " [*]" toe te voegen. Zo wordt bijvoorbeeld

```
Australia:1411:55.55:Mackay Cyclone 1918:1918:Mackay  
omgezet naar
```

```
Australia:1411:55.55:Mackay Cyclone 1918:1918:Mackay [*]
```

Probeer zo weinig mogelijk commando's te geven en zo weinig mogelijk tekens te gebruiken in je commando's. De commentaarregels mogen door je substituties niet gewijzigd worden. Alle wijzigingen moeten na elkaar uitgevoerd worden.

Opgave 3

Vanaf de zon geteld is Mars de vierde planeet van ons zonnestelsel, en ze draait rond de zon in een baan tussen die van de Aarde en Jupiter. De planeet is kleiner dan de Aarde en met een maximale magnitude van -2,9 is ze minder helder dan Venus en meestal minder helder dan Jupiter. De planeet is vernoemd naar de Romeinse god van de oorlog, al wordt ze in de volksmond ook wel de *rode planeet* genoemd terwijl ze in werkelijkheid eerder okerkleurig is. De marssondes die sinds 1964 werden gelanceerd binnen het Amerikaanse Mariner-programma slaagden erin systematisch foto's te nemen van de grote en kleine kraters op Mars. Hierop was te zien dat Mars bezaaid is met honderdduizenden kraters, waaronder enkele van de grootste van ons zonnestelsel. De kraters op Mars worden vernoemd naar aardse steden met minder dan 100.000 inwoners, geleerden en (science-fiction) schrijvers.

Het bestand `kraters.txt` bevat de officiële lijst met alle benoemde kraters op Mars, uitgegeven door de Amerikaanse U.S. Geological Service (USGS). Naast de naam van elke krater (eerste veld) bevat de lijst ook nog informatie velden met de coördinaten (`ligging`), de maximale diameter uitgedrukt in kilometer (`diameter`) en het jaar waarin de naam van de krater officieel werd goedgekeurd door de International Astronomical Union (`goedkeuring`). Met betrekking tot de herkomst van de naam bevat de lijst ook een korte omschrijving (`ethymologie`), samen met het continent (`continent`) en het land (`land`) waarin de naam zijn oorsprong vindt. De coördinaten bevatten de breedteligging uitgedrukt in noorderbreedte (N) of zuiderbreedte (Z) en de lengteligging uitgedrukt in westerlengte (W) of oosterlengte (O), van elkaar gescheiden door een koppelteken (-). Velden worden van elkaar gescheiden door tabs.

Gevraagd wordt om een `bash` shell script `statistieken.sh` te schrijven dat de volgende statistieken berekent voor het bestand `kraters.txt`:

- totaal aantal benoemde kraters
- aantal verschillende continenten
- aantal verschillende landen
- minimaal aantal kraters per continent
- maximaal aantal kraters per continent
- minimaal aantal landen per continent
- maximaal aantal landen per continent
- lijst van maximale kraterdiameters per continent

De naam van het bestand waarop de statistieken worden berekend, moet als parameter aan het shell script meegegeven worden. In het shell script mag je gebruik maken van IO-redirectie, pipes en filters, maar mag je geen gebruik maken `awk` (of de GNU versie daarvan). Indien het shell script wordt aangeroepen als

```
> statistieken.sh kraters.txt
```

dan moet het de volgende informatie wegschrijven naar standaard uitvoer:

```
aantal records:      958
aantal continenten:  xxx
aantal landen:       xxx
min kraters/continent: xxx
max kraters/continent: xxx
min landen/continent: xxx
max landen/continent: xxx
max diameter per continent:
  Africa              185
  Asia                 xxx
  Europe               xxx
  ...
```

Op de plaats van de xxx komen natuurlijk getallen te staan, overeenkomstig de gevraagde statistieken.

Opgave 4

1. Geef \LaTeX -code die een reconstructie van Listing 1 genereert, inclusief het onderschift. Zorg daarbij dat de opmaak zo getrouw mogelijk behouden blijft. Merk op dat het sleutelwoord “yield”, net zoals “def”, “for”, “in”, “if” en “import”, in het vet moet weergegeven worden.

```
1  def mapper(key, value):
2      for word in value.split():
3          yield word, 1
4
5  def reducer(key, values):
6      yield key, sum(values)
7
8  if __name__ == "__main__":
9      import dumbo
10     dumbo.run(mapper, reducer, combiner=reducer)
```

Listing 1: Python programma dat kan gebruikt worden om het aantal voorkomens van elk woord te tellen in een grote collectie tekstbestanden met behulp van Dumbo (<http://last.fm/dumbo>) en Hadoop (<http://hadoop.apache.org>).

2. Geef L^AT_EX-code die precies hetzelfde resultaat oplevert als het tekstfragment in onderstaand kader. Zorg ervoor dat formules en eigen omgevingen automatisch genummerd worden, en gebruik waar mogelijk verwijzingen naar deze nummeringen.

De basis voor de representaties bestaat uit alle Gelfand-Zetlin patronen voor partities met een maximum lengte n . Deze GZ-vectoren hebben de volgende vorm:

$$|m\rangle \equiv |m\rangle^n \equiv \left(\begin{array}{cccc} m_{1n} & \cdots & m_{n-1,n} & m_{nn} \\ m_{1,n-1} & \cdots & m_{n-1,n-1} & \\ \vdots & \ddots & & \\ m_{11} & & & \end{array} \right) \equiv \left(\begin{array}{c} [m]^n \\ [m]^{n-1} \end{array} \right). \quad (1)$$

Definitie 1 De **Hamiltoniaan** van een systeem in termen van $OSP(1|2n)$ generatoren wordt gegeven door de vergelijking (2):

$$\hat{H} = \hbar \sum_{j=1}^n \sqrt{c \lambda_j + \omega^2} h_j. \quad (2)$$

De vectoren van $|m\rangle$ zijn eigenvectoren van de Hamiltoniaan. We kunnen schrijven dat

$$\hat{H} |m\rangle = \hbar E_m |m\rangle,$$

waarbij E_m staat voor

$$E_m = \sum_{j=1}^n \sqrt{c \lambda_j + \omega^2} \left(\frac{p}{2} + \sum_{r=1}^j m_{rj} - \sum_{r=1}^{j-1} m_{r,j-1} \right). \quad (3)$$

De multipliciteit $\mu(E_k^{(p)})$ van elke eigenwaarde kan bepaald worden met behulp van enkele theoretische argumenten. Ten eerste, $\mu(E_k^{(p)})$ zal gelijk zijn aan het totale aantal GZ-vectoren met een partitie λ in de bovenste rij, waar λ een partitie van k in ten meeste p delen voorstelt. We weten dat de voorstelling van $GL(n)$ (die gelabeld is door de partitie λ) van dimensie $\binom{n}{\lambda'}$ is. Dit wordt gedefinieerd door

Definitie 2 (representatie van $GL(n)$)

$$\begin{aligned} \binom{X}{\lambda} &= \prod_{(i,j) \in \lambda} \frac{X - c(i,j)}{h(i,j)} \\ c(i,j) &= j - i \\ h(i,j) &= \lambda_i + \lambda'_j - i - j + 1 \end{aligned}$$

Voor een gegeven partitie λ is het aantal GZ-patronen dat λ in the bovenste rijen bevat gelijk aan $\binom{n}{\lambda}$. Dit heeft als gevolg dat de multipliciteit van elke eigenwaarde gelijk is aan

$$\mu(E_k^{(p)}) = \sum_{\lambda, |\lambda|=k, l(\lambda) \leq [p]} \binom{n}{\lambda'}$$

De ceiling functie $[p]$ wordt gebruikt om ook de gevallen waarbij $n - 1 < p < n$ te includeren.

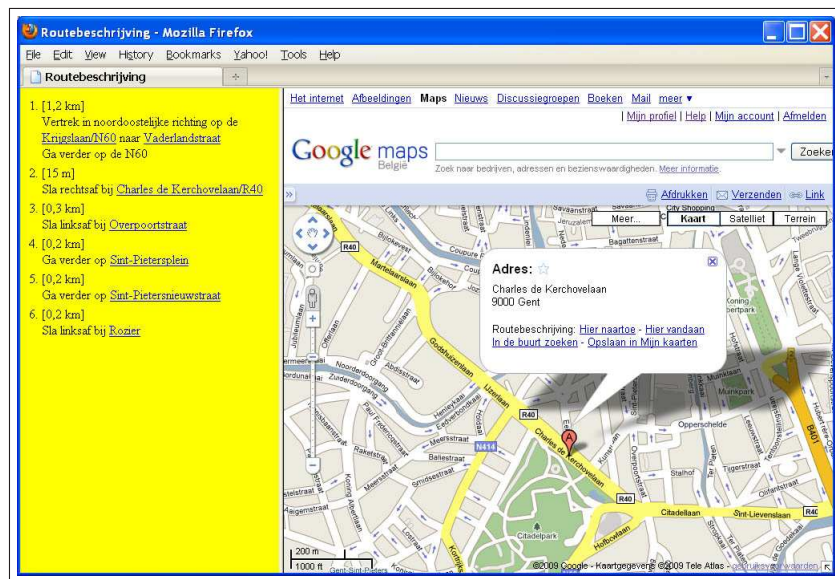
Plaats een PDF bestand met daarin de gecompileerde L^AT_EX fragmenten in je ZIP bestand.

Opgave 5

Gevraagd wordt om een `bash` shell script `bepaalroute` te schrijven dat de gebruiker toelaat om een routebeschrijving te genereren van een gegeven beginpunt naar een gegeven eindpunt. De adressen van het beginpunt en het eindpunt moeten als parameter aan het shell script doorgegeven worden. De routebeschrijving moet opgemaakt worden in XHTML 1.0 formaat. Daarbij moet de weergave gebruik maken van twee verticale frames, waarbij de omschrijving van de te volgen route in het linkerframe wordt weergegeven. Zo moet het commando

```
> bepaalroute "Krijgslaan 281, 9000 Gent" "Rozier 44, 9000 Gent"
```

een routeomschrijving genereren van Campus De Sterre naar Campus Rozier, die als volgt in een browser wordt weergegeven:



Bovendien moeten alle webpagina's die het script zelf genereert voldoen aan de XHTML 1.0 standaard. Om een routebeschrijving te genereren, maak je gebruik van de volgende Google Maps query

```
http://maps.google.be/maps?saddr=<beginpunt>&daddr=<eindpunt>
```

Binnen deze template-URL vul je op de plaats van `<beginpunt>` het adres van het beginpunt in en op de plaats van `<eindpunt>` het adres van het eindpunt. Let erop dat de omschrijving van het adres meestal uit meerdere woorden zal bestaan.

De gebruikte Google Maps query geeft een XHTML bestand terug, waaruit de nodige informatie kan gefilterd worden voor het opmaken van de XHTML bestanden die door het shell script `bepaalroute` moeten gegenereerd worden. Daarbij zit de routebeschrijving vervat in een XHTML tabel met als klasse-attribuut `ddr_steps` (je mag er vanuit gaan dat het XHTML dat Google Maps teruggeeft slechts één tabel met een dergelijk klasse-attribuut bevat). Elke rij uit deze tabel correspondeert met één enkele stap uit de te volgen route en bestaat uit drie kolommen die respectievelijk het stapnummer, een omschrijving van de route en de lengte van deze stap uit de route aangeven. Het onderstaande tekstfragment geeft aan hoe de route uit bovenstaand voorbeeld wordt omschreven in het XHTML document dat door Google Maps wordt teruggegeven. Let hierbij op het feit dat bijkomende regels werden toegevoegd om de weergave van het fragment te verbeteren, daar waar het document dat Google Maps teruggeeft deze tabel op één enkele regel bevat.

```
<table class=ddr_steps id=ddr_steps_0>
  <tr class="dirsegment firststep" id=step_0_0>
```

```

<td class=num>1.</td>
<td class=dirsegtext id=dirsegtext_0_0>Vertrek in <b>noordoostelijke</b> richting op de
<b>Krijgslaan/<wbr/>N60</b> naar <b>Vaderlandstraat</b>
<div class="dirsegnote note_CONTINUE">Ga verder op de N60</div></td>
<td class=sdist><div id=sxdist class=nw>1,2<#160;km</div></td>
</tr>
<tr class=dirsegment id=step_0_1>
<td class=num>2.</td>
<td class=dirsegtext id=dirsegtext_0_1>Sla <b>rechtsaf</b> bij <b>Charles de
Kerchovelaan/<wbr/>R40</b></td>
<td class=sdist><div id=sxdist class=nw>15<#160;m</div></td>
</tr>
<tr class=dirsegment id=step_0_2>
<td class=num>3.</td>
<td class=dirsegtext id=dirsegtext_0_2>Sla <b>linksaf</b> bij <b>Overpoortstraat</b></td>
<td class=sdist><div id=sxdist class=nw>0,3<#160;km</div></td>
</tr>
<tr class=dirsegment id=step_0_3>
<td class=num>4.</td>
<td class=dirsegtext id=dirsegtext_0_3>Ga verder op <b>Sint-Pietersplein</b></td>
<td class=sdist><div id=sxdist class=nw>0,2<#160;km</div></td>
</tr>
<tr class=dirsegment id=step_0_4>
<td class=num>5.</td>
<td class=dirsegtext id=dirsegtext_0_4>Ga verder op <b>Sint-Pietersnieuwstraat</b></td>
<td class=sdist><div id=sxdist class=nw>0,2<#160;km</div></td>
</tr>
<tr class=dirsegment id=step_0_5>
<td class=num>6.</td>
<td class=dirsegtext id=dirsegtext_0_5>Sla <b>linksaf</b> bij <b>Rozier</b></td>
<td class=sdist><div id=sxdist class=nw>0,2<#160;km</div></td>
</tr>
</table>

```

In het linkerframe maak je een tabel op met in de eerste kolom het nummer van de stap uit de routebeschrijving gevuld door een punt. In de tweede kolom staat op de eerste regel de lengte van deze stap uit de route tussen vierkante haakjes, gevolgd door een omschrijving van de route op de volgende regel. Indien de omschrijving een `div`-element bevat van de klasse `dirsegnote`, dan moet de inhoud van het `div`-element worden weergegeven op een nieuwe regel (zie bijvoorbeeld de tekst *Ga verder op de N69* uit bovenstaand voorbeeld).

Indien de omschrijving een `b`-element bevat, dan mogen de bijhorende tags verwijderd worden indien de inhoud van het `b`-element start met een kleine letter. In het andere geval gaan we ervan uit dat de inhoud van het `b`-element een plaats- of straataanduiding weergeeft. Zorg ervoor dat als de gebruiker op een dergelijke plaats- of straataanduiding klikt, er een corresponderende webpagina wordt geopend in het rechterframe. De URL die correspondeert met een plaats- of straataanduiding heeft de vorm

`http://maps.google.be/maps?q=<aanduiding>`

waarbij op de plaats van `<aanduiding>` een plaats- of straataanduiding moet worden ingevuld. Zo correspondeert de plaatsaanduiding *Sint-Pietersplein* met de volgende URL

`http://maps.google.be/maps?q=Sint-Pietersplein`

Het rechterframe mag leeg zijn zolang de gebruiker nog geen link in het linkerframe heeft aangeklikt.

Opmerking: Indien je er niet in slaagt om een gepast XHTML document terug te krijgen van Google Maps, dan kun je gebruik maken van het meegeleverde bestand `googlemaps.html`. Dit bestand bevat het XHTML document dat normaalgezien door Google Maps wordt teruggegeven als gezocht wordt op de route van *Krijgslaan 281, 9000 Gent* naar *Rozier 44, 9000 Gent*.