

Generiese Vaardighede: gawk oefening

Abrie Greeff
B.Sc Hons (Rekenaarwetenskap)
Departement Rekenaarwetenskap
Universiteit van Stellenbosch

24 April 2006

Part I

Genomedata

Vraag 1

Soek vir die patroon GGTTAA en vertoon die DE lyn van die groepe waar die patroon gevind is.

Om dit te doen moet die volgende uitgevoer word.

```
gawk -f ../q0.awk genomedata.txt
```

Die awk script lyk soos volg:

```
BEGIN {n = 0; id = ""; de = "";}
{
  SY = $9
  id = $1
  if (id ~ /DE/) {
    de = $0
  }
  if (SY == "") {
    n = 0;
  }
  if (SY ~ /'G';/) {
```

```
    if ((n != 0) && (n != 1))
      n = 0
    else
      if (n == 1)
        n = 2
      else
        if (n == 0)
          n = 1
```

```
  }
  if (SY ~ /'T';/) {
    if ((n != 2) && (n != 3))
      n = 0
    else
      if (n == 3)
        n = 4
      else
        if (n == 2)
```

```

n = 3

}
if (SY ~ /'A';/) {
if ((n != 4) && (n != 5))
n = 0
else
if (n == 5)
n = 0
print de
else
if (n == 4)
n = 5

}
if (SY ~ /'C';/) {
n = 0
}
}

```

Part II

Test data

Vraag 2

Orals waar "start" gevind word vervang dit met "START".

Om dit te doen moet die volgende uitgevoer word.

```
ls | gawk -f ../q1.awk | csh
```

Die *directory* waar dit uitgevoer word is een vlak laer as waar die awk script lê. Die awk script lyk soos volg:

```

{for (i = 0;i < NF;i++) {
print "sed -e 's/start/START/g' -i " $i
}}

```

Vraag 3

Tel hoeveel kere die woord "FINAL" voorkom.

Om dit te doen moet die volgende uitgevoer word.

```
cat * | gawk -f ../q2.awk
```

Die awk script lyk soos volg:

```
BEGIN {count = 0}  
/FINAL/ {count = count + 1}  
END {print count}
```

Vraag 4

Tel hoeveel files bevat minder as vyf lyne teks.

Om dit te doen moet die volgende uitgevoer word.

```
ls | gawk -f ../q3.awk | csh | gawk -f q2.awk
```

Die awk scripts lyk soos volg:

```
# q3.awk  
{for (i = 0;i < NF;i++) {  
  print "gawk -f q3_2.awk " $i  
}}
```

en

```
# q3_2.awk  
END { if (NR < 5) print NR "FINAL" }
```

Vraag 5

Verander al die toets files sodat die lyne wat die woord "FINAL" bevat voorkom net na die lyn wat "START" bevat.

Om dit te doen moet die volgende uitgevoer word.

```
ls | gawk -f ../q4.awk | csh
```

Die awk scripts lyk soos volg:

```
# q4.awk  
{for (i = 0;i < NF;i++) {  
  print "echo Changing " $i  
  print "gawk -f ../q4_2.awk " $i " > temp"  
  print "cat temp > " $i  
}  
  print "rm -f temp"}
```

en

```
# q4_2.awk
{n=0}
/START/ {teks = $0; n = 1}
/FINAL/ {teks = teks "\n" $0; n = 1}
n==0 {teks2 = teks2 "\n" $0;n=0}
END {print teks,teks2}
```