Universität Ulm - Abteilung Angewandte Informationsverarbeitung

Dr. Johannes Mayer und Hans Braxmeier, Wintersemester 2003

10. Übungsblatt zur Vorlesung Allgemeine Informatik III

Abgabetermin: Dienstag, 20.01.2004

Benford's Law *(10 Punkte)*

Ausschlaggebend für die Entdeckung von Benford's Law waren Logarithmentafeln, die auf den vorderen Seiten stärker abgegriffen waren, als auf den hinteren Seiten, so als hätten sich die Benutzer bevorzugt für die niedrigen Zahlen interessiert...

Werfen wir einen Blick in eine Tageszeitung treffen wir auf Zahlen aller Art: Benzinpreise, Einkommen von Familien, Unfallstatistiken, Hochwasserstände, Einwohnerzahlen von Städten usw. Soll nun die Wahrscheinlichkeit, daß eine beliebig herausgegriffene Zahl mit einer 1 beginnt geschätzt werden, so erhält man meist 1/9 als Antwort. Die meisten Leute gehen also von einer Gleichverteilung der Anfangsziffern aus (mal abgesehen von führenden Nullen).

Doch diese Annahme ist falsch. Analysiert man z.B. die Anfangsziffern von Bevölkerungsdichten, Atomgewichten, Luftdruckmessungen usw. stellt man fest: Die 1 als Anfangsziffer liegt immer vorn! In etwa 1/3 aller Fälle fängt die Zahl mit einer 1 an, nur in etwa 1/22 der Fälle mit einer 9!

Testen Sie Benford's Law!

Testen Sie Benford's Law, indem Sie die Größe und die Zeitstempel von Dateien in verschiedenen Verzeichnissen auf Ihrer Festplatte analysieren.

a) Schreiben Sie hierfür eine Prozedur, die ein Verzeichnis und alle seine Unterverzeichnisse rekursiv abläuft. Das Startverzeichnis soll als Kommandozeilenargument dem Programm übergeben werden. (*4 Punkte*)

b) Erweitern Sie Ihr Programm so, daß die Größe und die Zeitstempel aller darin enthaltenen Dateien ausgelesen werden. (*3 Punkte*)

c) Stellen Sie anschließend die Häufigkeitsverteilungen der Anfangsziffern von Größe und Zeitstempel tabellarisch dar. (*3 Punkte*)

d) Nur für Mathematiker: Versuchen das Resultat zu deuten... (*Fields-Medaille, wer die Erklärung von Hill (1996) nachvollziehen kann* ;-)

Anhand von Beispielprogrammen wird in den Übungen der Zugriff auf das Dateisystem von Unix erklärt. Zwei Bemerkungen hierzu: Es genügt, den Zeitstempel als Integer-Zahl zu lesen (eine Umrechnung in Stunden und Minuten ist nicht verlangt) und verwenden sie zum Lesen von Dateiinformationen (INode) den Befehl `lstat` anstelle von `stat`.

Beispielausgabe

```
hb1@thales:/home/hb1> a.out /home/hb1/temp

           Dateiname |      Dateigroesse     |       Zeitstempel

       Elements.java |               123 (1) |        1070272499 (1)
          Point.java |               456 (4) |        1070272499 (1)
         Tester.java |               424 (4) |        1070272499 (1)
        Tester.class |               721 (7) |        1070272544 (1)
CANNOT READ INFO OF /home/hb1/Window$1.class
          Line.class |               889 (8) |        1070272449 (1)
           Line.java |               574 (5) |        1070272439 (1)
         Point.class |               806 (8) |        1070272449 (1)
          Point.java |               312 (3) |        1070272439 (1)
         Tester.java |               334 (3) |        1070272439 (1)
        Tester.class |               607 (6) |        1070272449 (1)
           java.prof |             93633 (9) |        1070044392 (1)
        HPjmeter.jar |            952803 (9) |        1043861965 (1)
       MyCanvas.java |              1742 (1) |        1073511790 (1)
CANNOT OPEN DIR /home/hb1/temp/geheim
         Start.class |               374 (3) |        1073511796 (1)
                 ... |               ...     |              ...

Häufigkeitsverteilungen der Anfangsziffern von Dateigröße und Zeitstempel
bei der Analyse von 30135 Dateien:

size[Ziffer 1]:    9020   |   time[Ziffer 1]   27654
size[Ziffer 2]:    5793   |   time[Ziffer 2]       0
size[Ziffer 3]:    4136   |   time[Ziffer 3]       0
size[Ziffer 4]:    3269   |   time[Ziffer 4]       0
size[Ziffer 5]:    1819   |   time[Ziffer 5]       0
size[Ziffer 6]:    1909   |   time[Ziffer 6]       4
size[Ziffer 7]:    1472   |   time[Ziffer 7]      33
size[Ziffer 8]:    1573   |   time[Ziffer 8]     117
size[Ziffer 9]:    1045   |   time[Ziffer 9]    2327
```

Eine Erklärung zu Benford's Law oder wie das Gesetz zur Überprüfung von Steuererklärungen herangezogen werden kann, beschreibt der Artikel „Die Welt vom Planeten Zob“ von Jörg Albrecht, erschienen in DIE WELT:

`http://www.mathematik.uni-karlsruhe.de/Presse/eins_bearb.html`.

Viel Erfolg!