Dr. Andreas Borchert
Ingo Melzer

Implementierung kleiner Datenbanken unter UNIX II (WS 1999/2000)

Abgabetermin 18. November 1999

4 Jeder kennt jeden (20 Punkte)

Vor einigen Jahren war die Suche nach Schauspielern mit einer hohen Kevin Bacon Nummer beliebt. Dabei geht es um die Suche nach dem kürzesten Weg zwischen zwei Personen, wobei die Anzahl der Sprünge über andere Personen gezählt werden. Wenn also A B kennt und B C kennt, so ist der Abstand zwischen A und C zwei. Um dieses Spiel zu spielen, haben wir entsprechende Daten zu Anna Karenina von Leo Tolstoy, David Copperfield von Charles Dickens, Homer, Huckleberry Finn von Mark Twain und Les Misérables von Victor Hugo auf unserem FTP-Server gelegt. Diese stammen von der Stanford GraphBase von Donald E. Knuth.

Schreiben Sie nun ein Perl Skript, das zuerst eine dieser Dateien einliest, dann ein Kürzel einer beteiligten Person bekommt und dann schrittweise die transitive Hülle dieser Person bestimmt (und natürlich ausgibt). Dies könnte z. B. so aussehen:

```
thales$ ./related.pl huck.dat
The following persons have been found: AB, AP, AS, AT, BD, BE, BG, BH,
BI, BK, BM, BN, BO, BP, BR, BS, BT, BU, CG, CS, DH, DR, DU, HF, HI, HS,
HT, HW, JG, JH, JI, JK, JL, JM, JN, JO, JP, JT, JY, KI, LB, LI, LZ, MA,
MC, MH, MJ, MP, MR, MS, MW, NT, OD, PA, PW, RG, RH, SD, SI, SP, SR, SS,
SU, SW, TB, TC, TF, TG, TP, TS, TU, WB, WD, WW.
Please enter the person's key: AB
AB knows in 0 steps: AB
AB knows in 1 steps: BN DH DR DU HF JO KI MJ PW SW TF WB
AB knows in 2 steps: AP AS AT BD BE BG BH BI BK BM BO BP BR BS BT BU CG
CS HI HS HT HW JG JH JI JK JL JM JN JP JT JY LB LI LZ MA MC MH MP MR MS
MW NT OD PA RG RH SD SI SP SR SS SU TB TC TG TP TS TU WD WW
thales$
```

Noch ein paar Tips:

- Achten Sie auf das Format der Eingabedateien. Ein & bedeutet z. B., daß die aktuelle Zeile nur die Fortsetzung der letzten ist.
- Als Datenstruktur ist ein Hash von Hashes zu empfehlen. Hierfür sind Zeiger zu verwenden.
- Achten Sie darauf, daß jede Person nur einmal ausgegeben wird und daß keine Zyklen entstehen.
- Die Aufgabe besteht eigentlich aus drei Teilen. Dem Einlesen der Daten, der Lagerung in einer sinnvollen Datenstruktur und zum Schluß der Ausgabe in der gewünschten Form.