Prof. Dr. Franz Schweiggert und Dr. Johannes Mayer Abt. Angewandte Informationsverarbeitung
19. Mai 2006
Ralph Guderlei Blatt 4

Software Engineering Praxis (SS 2006)

Abgabetermin: 26. Mai 2006

1 Erdös-Zahl (10 Punkte)

Paul Erdöes ist einer der bekanntesten Mathematiker des 20. Jahrhunderts. Er publizierte mit einer großen Zahl anderer Autoren, was ihn auf die Idee brachte, die Erdös-Zahlen wie folgt zu definieren:

- Paul Erdös hat Erdös-Zahl 0
- Jeder Koautor von Paul Erdös hat Erdös-Zahl 1
- Jeder Koautor eines Koautors von Erdös hat Erdös-Zahl 2
- usw.

Die Erdös-Zahl ist eine Veranschaulichung des Kleine-Welt-Phänomens von Stanley Milgram, der behauptet, dass in einer modernen Gesellschaft alle über eine unerwartet kleine Anzahl an persönlichen Bekanntschafen zueinander in Verbindung stehen.

Die Erdös-Zahl kann über einen Beziehungs-Graphen berechnet werden. Die Autoren sind die Knoten des Graphen und zwischen zwei Knoten besteht eine Kante, wenn zwei Autoren miteinander publiziert haben.

Der Graph kann also aus einer Literatur-Datenbank aufgebaut werden. Eine *kleine Datenbank* steht zur Erzeugung zur Verfügung. Um die Berarbeitung der Datenbank zu vereinfachen, steht ein *jar-file* zur Verfügung, welches Methoden zur Zerlegung der Einträge zur Verfügung stellt. Schreiben Sie also als erstes ein Programm zum Einlesen der Datenbank und speichern Sie die Autoren und den Titel der Veröffentlichungen ab. Jede Zeile der Datenbank-Datei entspricht einem Eintrag, die Autoren und den Titel erhalten Sie dann mit Hilfe der im jar-file zur Verfügung gestellten Klasse.

Als nächstes schreiben Sie ein Programm, welches aus den Daten die Erdös-Zahl jedes Autors bestimmt. Stellen Sie hierzu den Beziehungs-Graphen auf. Sie können davon ausgehen, dass jeder Nachname eindeutig einen Autoren bestimmt. Die Erdös-Zahl eines Autors X lässt sich nun als Länge des kürzesten Wegs im Graphen von X zu Paul Erdös berechnen, wobei jede Kante die Länge 1 hat.

Um den Graphen aufzubauen reicht es aus, in einer geeigneten Datenstruktur für jeden Autor dessen Koautoren abzulegen. Beispielsweise eignet sich eine Map. Der Schlüssel ist der Name des Autors, der abgelegte Wert ist die Liste der Koautoren.

Zur Bestimmung aller Pfade zwischen zwei Knoten können zwei Algorithmen, die Tiefen- oder Breitensuche verwendet werden. In beiden Verfahren werden ausgehend von einem Start-Knoten alle verbundenen Knoten besucht bis der gesuchte Knoten gefunden wurde. Folgender Code beschreibt das Vorgehen um einen krzesten Pfad zum Knoten ”ziel” zu suchen:

```
class Node{
        public int number;
        public int dist;
        public Node(int n, int d){
                number = n ; dist  = d;
        }
        public boolean equals(Object o){
                if(!(o instanceof Node)) return false;
                return ((Node)o).number == number && ((Node)o).dist == dist;
}

public int bfs(Node start, Node ziel) {
        Queue<Node> toBeVisited = new LinkedList<Node>();
        Set<Node> visited = new TreeSet<Node>();

        toBeVistited.add(start);
        while (!toBeVisited.isEmpty()) {
                Node node = toBeVisited.poll();
                if (node.equals(ziel)
                        return node.dist;
                visited.add(node);
                for (Node neighbor : getNeighbors(node))
                        if (!visited.contains(neighbor))
                                toBeVisited.add(new Node (neighbor.number, node.dist+1}
        }

        return -1; // nicht erreichbar
}
```

section*Hinweise

- *Wikipedia: Breitensuche*
- *Erdös Number Project*
- *Wikipedia: Kleine-Welt-Phänomen*

Viel Erfolg!