

Präsenzblatt 1

Dieses Blatt dient lediglich der persönlichen Vorbereitung. Es wird nicht abgegeben und geht nicht in die Bewertung ein. Die Besprechung der Aufgaben und ihrer Lösungen erfolgt in den kleinen Übungen in der Woche vom 18.11.2024.

Präsenzaufgabe:

Dem *Orakel von Kevin Bacon* liegt der Schauspielergraph S zugrunde: Schauspieler sind durch Knoten repräsentiert. Zwei Schauspielerknoten sind durch eine Kante verbunden, wenn sie gemeinsam in einem Film gespielt haben. Der Knoten von Kevin Bacon hat die *Kevin-Bacon-Zahl* (KBZ) 0; die KBZ eines anderen Schauspielers ist die Länge eines kürzesten Weges im Schauspielergraphen S zu Kevin Bacon. Beispielsweise hat Tom Hanks die KBZ 1, da er mit Kevin Bacon in *Apollo 13* gespielt hat. Falls kein verbindender Weg existiert, ist die KBZ des betrachteten Schauspielers als unendlich definiert.

Das Orakel ist im Web verfügbar: <http://oracleofbacon.org/>. Die zugrundeliegenden Filmdata sind von <https://www.themoviedb.org/> entnommen.

- a) Wir betrachten einen Pfad für die KBZ z eines Schauspielers A als ein kürzester Pfad in S , der Kevin Bacon und A verbindet und der aus $z + 1$ Knoten besteht. Gib einen Schauspieler mit mindestens KBZ 4 und einen entsprechenden Pfad an.
- b) Wenn man Schauspieler mit möglichst großer KBZ sucht, muss man davon ausgehen, dass man diese nicht kennt; man muss sie also erst als Teil der Suche finden. Beschreibe eine allgemeine Strategie zum sicheren Finden (ohne Raten) von Schauspielern mit hoher KBZ, die jeweils nur ein Browserfenster/-tab mit dem Orakel von Bacon und ein Browserfenster/-tab mit TMDb verwendet. Dass man für diese Strategie in der Praxis sehr viel Zeit mitbringen müsste, ist in Ordnung. Welche Rolle spielt bei dieser Strategie die Breitensuche? Welche spielt die Tiefensuche?
- c) Wenn man einen Server wie das Kevin-Bacon-Orakel betreibt, muss man damit rechnen, dass in kurzer Zeit sehr viele Anfragen hereinkommen, die jeweils schnell beantwortet werden müssen. Deshalb lohnt es sich, zwischen Verfahren zu unterscheiden, die eine Aufgabenstellung nur einmal lösen, und solchen, die (nach einem gewissen Aufwand für "Preprocessing" zur Erstellung einer geeigneten Datenstruktur) dieselbe Frage immer wieder neu für unterschiedliche Anfragen ("Queries") beantworten können. Bei einer Anfrage soll nun nicht nur die KBZ des betrachteten Schauspielers, sondern auch ein entsprechender Pfad ausgegeben werden, siehe Aufgabenteil a). Wie kann man es als Betreiber eines Orakels vermeiden, dass man für jede Anfrage eine neue Breitensuche ausführen muss, ohne dass man gigantische Datenmengen vorhalten muss?¹

¹Hier geht es nicht um rein technische Lösungen wie Caching, sondern um das Ausnutzen von Struktur.