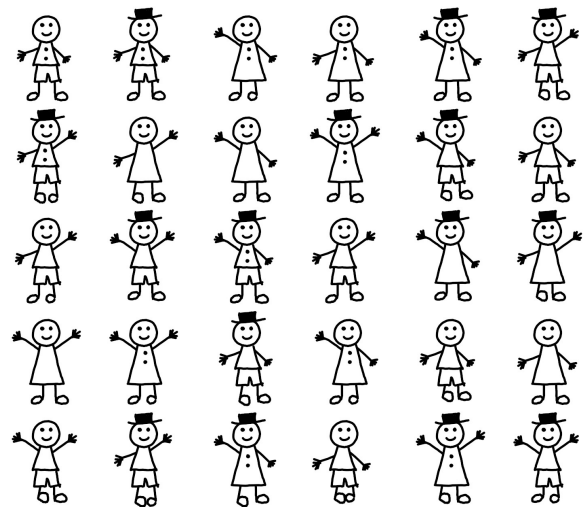


Das doppelte Lottchen: Suchen und Sortieren

© Michael Eisermann, Stefan Kohl, Friederike Stoll

Sie machen ein Praktikum bei einer Versicherung. Jeder der 10 Millionen Kunden hat eine 15-stellige Kundennummer und weitere Versicherungsdaten, die in einer Datenbank abgespeichert sind. Diese Datenbank hat vorerst noch keine besondere Struktur, insbesondere ist sie (noch) nicht nach Kundennummern sortiert. Sie sollen überprüfen, ob alle Kundennummern wirklich verschieden sind. Dazu verwenden Sie einen Computer mit Ihrer Lieblingsprogrammiersprache. Ein Vergleich ($<$, $=$, $>$) kostet etwa 1 Mikrosekunde, sonstige Operationen sind vernachlässigbar (Indexrechnung, nachschlagen, umordnen, etc). Beispielsweise benötigt das Auffinden der kleinsten Kundennummer 9 999 999 Vergleiche und dauert etwa 10 Sekunden.



Wo ist das Doppelgängerpaar?

Wenn Sie jedes Paar auf Dopplung prüfen, wie lange benötigt der Computer?

Wie lange dauert die Prüfung auf Dopplungen, wenn die Datenbank sortiert vorliegt?

Leider ist die Datenbank völlig unsortiert. Ihre Chefin möchte dies ändern und betraut Sie mit einer möglichst effizienten Sortierung. Diese soll ebenfalls nur paarweise Vergleiche nutzen und selbst im ungünstigsten Fall möglichst schnell sein. Sie gibt Ihnen noch den Tipp, dass Sie in der Literatur oder im Internet nach bekannten vergleichsbasierten Verfahren schauen sollten.

Wie lange benötigt der Computer mit einem solchen Verfahren zum Sortieren?

Beantworten Sie jede der drei Fragen mit einem der folgenden Zeitintervalle:

unter 1 Sek.	1 bis 5 Sek.	5 bis 20 Sek.	20 bis 60 Sek.	1 bis 3 Min.	3 bis 5 Min.	5 bis 60 Min.
1 bis 24 Stunden	1 bis 7 Tage	1 bis 4 Wochen	1 bis 12 Monate	1 bis 3 Jahre	3 bis 10 Jahre	länger als 10 Jahre