

Ideen der Informatik

Suchen und Sortieren

[Ordnung muss sein...]

Kurt Mehlhorn
Adrian Neumann

viele Folien von Kostas Panagiotou

Suchen

- Welche Telefonnummer hat Kurt Mehlhorn?
- Wie schreibt man das Wort “Äquivalenz”?
- Welche Webseiten enthalten die W6rter “Uni Saarland”? Welche ist die “wichtigste”?

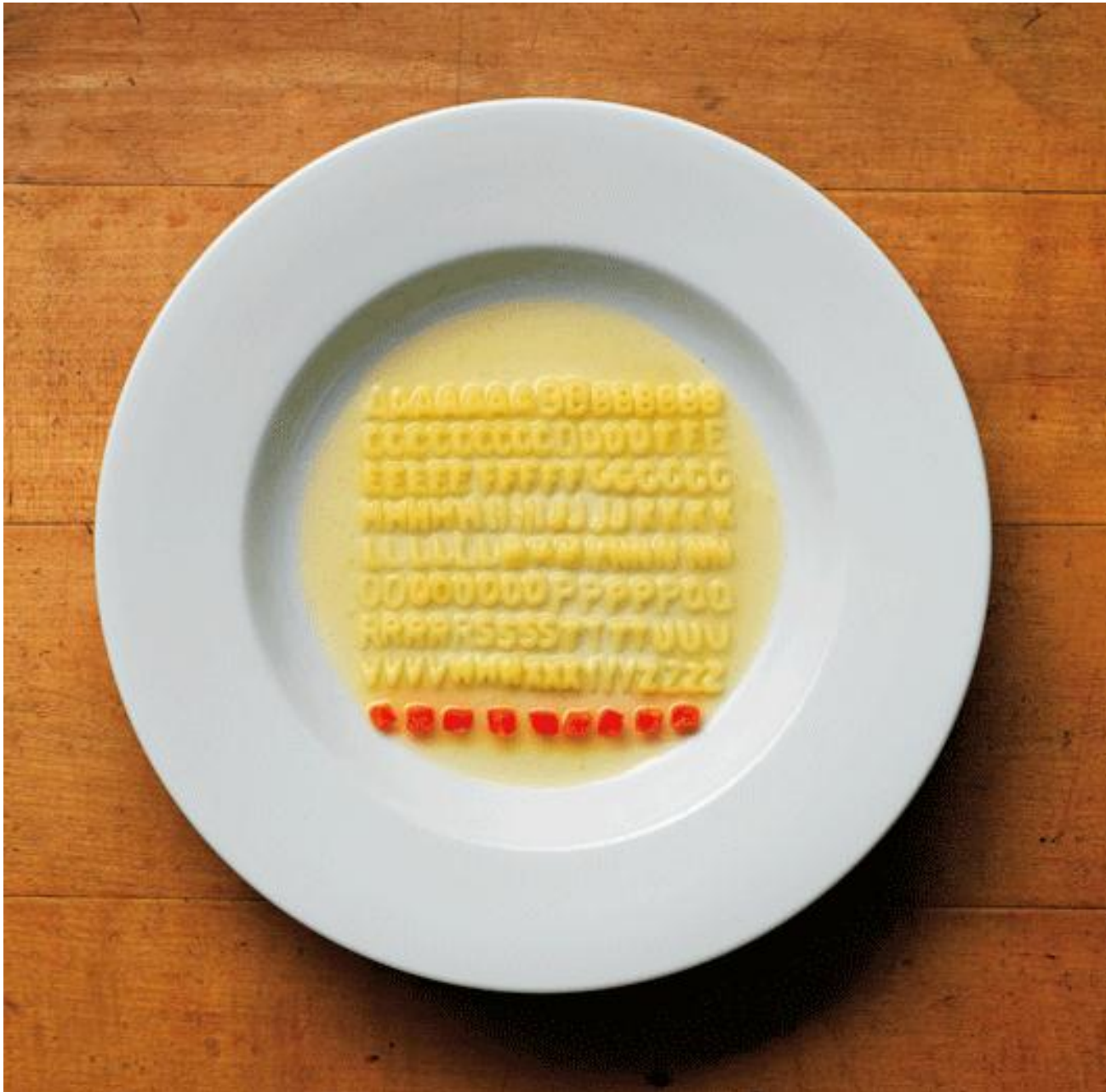
Ordnung erleichtert
das Suchen



Gibt es ein X in der Buchstabensuppe?



Aha!



Ein grünes Badetuch?



Aha!



*„Freibad heißt ja nicht, dass es jedem freisteht, einfach irgendwo rumzuliegen“
Ursus Wehrli, 2011*

Konzepte der heutigen Vorlesung

- Suchen = Information ablegen und verarbeiten, so dass man sie schnell wiederfindet (oder sagen kann, dass sie nicht vorhanden ist)
- Ordnung erleichtert das Suchen
- Sortieren = nach einem Kriterium ordnen
- Datenstrukturen = Suchen in Mengen, die sich zeitlich ändern

Suchen

- Daten können alles Mögliche sein:
 - Zeichenketten, Zahlen, Bilder, ...
- Hier: (Name + Telefonnummer)
- Haben einen Karteikasten: auf jeder Karteikarte steht ein Name und eine Nummer
- Wieviele Zettel muss man anschauen, bis man die Nummer zu einem Namen hat?

Zettel sind ungeordnet

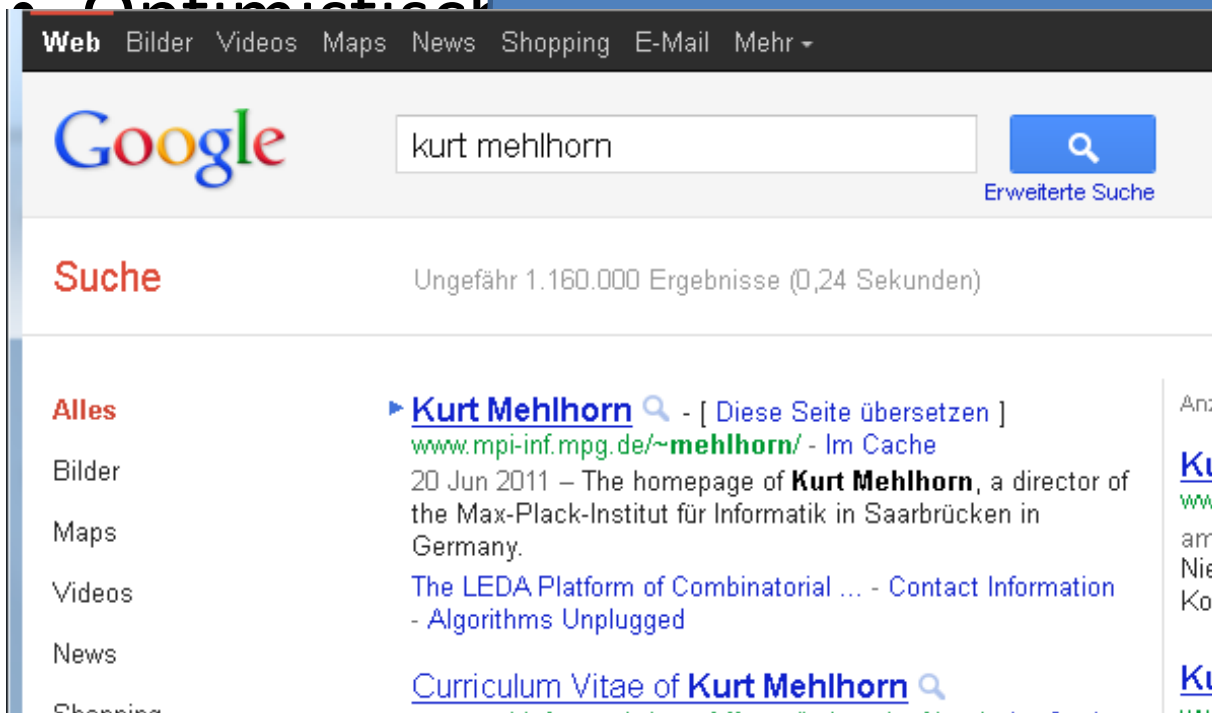
- Wir müssen **alle Karten** anschauen, um sicher zu sein, dass ein gesuchter Name nicht da ist.
- Falls ein Name da ist, im Mittel die Hälfte der Karten

Anzahl der Vergleiche = Anzahl der Karten

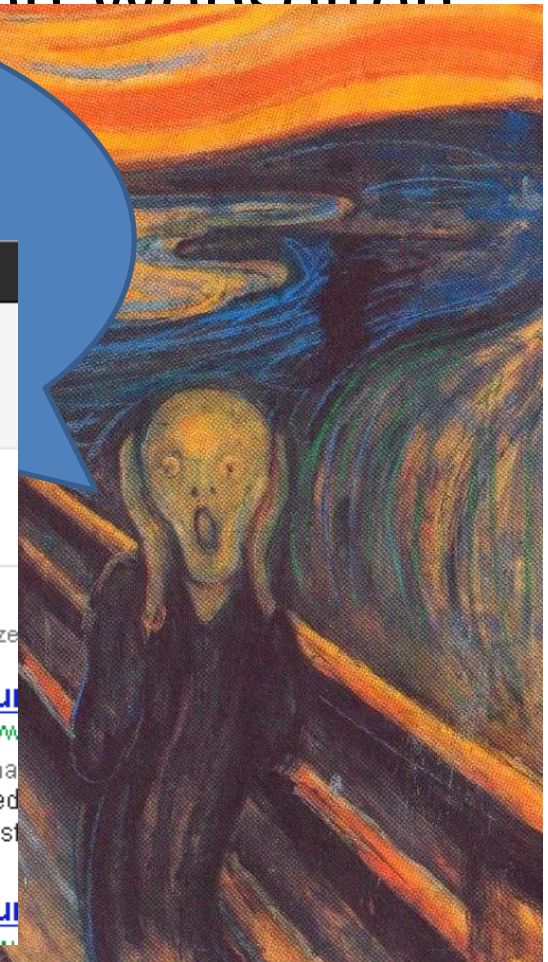
Ein Beispiel

- Das Internet hat mehrere Billionen Webseiten
- 1 Billion = 1.000.000.000.000
- Optimistisch

Ich finde Kurt's



The screenshot shows a Google search interface. At the top, there are navigation links: Web, Bilder, Videos, Maps, News, Shopping, E-Mail, Mehr. The Google logo is on the left, and the search bar contains the text "kurt mehlhorn". To the right of the search bar is a blue search button with a magnifying glass icon and the text "Erweiterte Suche". Below the search bar, the word "Suche" is displayed in red, followed by the text "Ungefähr 1.160.000 Ergebnisse (0,24 Sekunden)". On the left side, there is a vertical menu with links: Alles, Bilder, Maps, Videos, News, Shopping. The main search results area shows a list of results. The first result is a blue link "Kurt Mehlhorn" with a magnifying glass icon, followed by a text snippet: "[Diese Seite übersetzen] www.mpi-inf.mpg.de/~mehlhorn/ - Im Cache". Below this is a date "20 Jun 2011" and a description: "The homepage of Kurt Mehlhorn, a director of the Max-Planck-Institut für Informatik in Saarbrücken in Germany." There are two more links: "The LEDA Platform of Combinatorial ... - Contact Information - Algorithms Unplugged" and "Curriculum Vitae of Kurt Mehlhorn".



Wie machen wir das besser?

- Wir sortieren unsere Karteikarten nach Name
 - Also wie in einem Telefonbuch
- Wir suchen nach X
- Wir ziehen eine Karte, darauf steht Y
- X kommt vor/nach Y in der alphabetischen Ordnung der Namen oder $X = Y$
- Was wissen wir nun? Welche Karte nehmen wir für den ersten Vergleich? Wie geht es weiter?

Binärsuche

- Gegeben:
 - Liste mit N Elementen
 - Sortiert: der Nachfolger ist
Vorgaenger.
- Frage: enthaelt die Liste ein Element x ?
- Algorithmus:

Konzept: Divide and Conquer



Komplexität

- $n = 1$: Kosten = 1
 - $n = 3$: Kosten = 2 $3 = 2 \times 1 + 1$
 - $n = 7$: Kosten = 3 $7 = 2 \times 3 + 1$
 - $n = 15$: Kosten = 4 $15 = 2 \times 7 + 1$
 - $n = 31$: Kosten = 5
-
- $15 = 2 \times 2 \times 2 \times 2 - 1 = 2^4 - 1$ Kosten 4
 - $31 = 2 \times 2 \times 2 \times 2 \times 2 - 1 = 2^5 - 1$ Kosten 5
 - $N = 1 \text{ Billion} = 2^{40} - 1$ Kosten 40

Das ist irre!!!!

Lineare Suche vs. Binärsuche

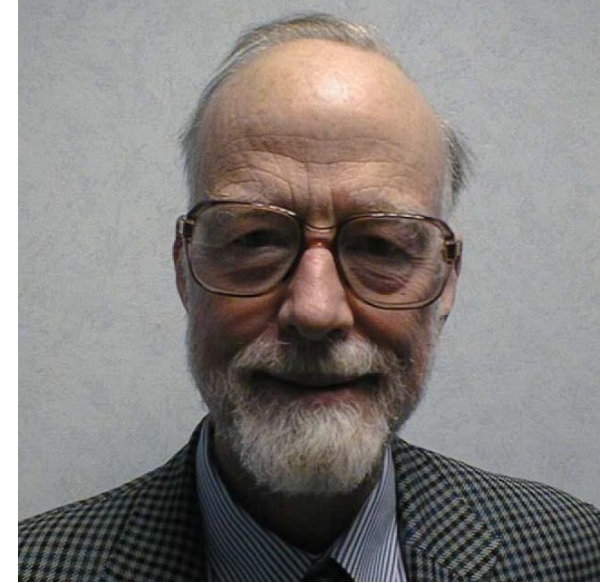
- Binärsuche funktioniert wenn man die gegebenen Daten ordnen kann:
 - (Name, Telefonnummer)
 - Webseiten?
- ***Lineare Suche: Aufwand = Anzahl der Elemente***
- ***Binärsuche: Aufwand = Logarithmus der Anzahl der Elemente***
- ***Binärsuche ist rasend schnell***

Wie sortiert man?

Quicksort

- S = Menge, die zu sortieren ist
- Wähle ein Element s in S
- Teile S in
 - $S_{<}$ = Elemente kleiner s
 - $S_{>}$ = Elemente größer s
- Gib aus

$\text{Sort}(S_{<}) \ s \ \text{Sort}(S_{>})$



Tony Hoare

Tony Hoare (1934 --

„Ich stelle fest, dass es zwei Wege gibt, ein Software-Design zu erstellen, entweder so einfach, dass es offensichtlich keine Schwächen hat, oder so kompliziert, dass es keine offensichtlichen Schwächen hat. Die erste Methode ist weitaus schwieriger.“

Tony Hoare, Dankesrede für den Turingpreis 1980^[1]

„I think Quicksort is the only really interesting algorithm that I've ever developed. “

Beim Teilen kann man Glück oder Pech haben

Laufzeit wie $n \log n$

n^2

$n = 10^6$ 0.1 sec

500 sec

Kann man das Glück erzwingen?

- Bis 1980: immer raffiniertere deterministische Strategien
- Seit 1980: wähle das Teilungselement zufällig
Randomisierter Algorithmus
- Urne mit $n/2$ roten und $n/2$ schwarzen Kugeln.
Wie findet man eine rote Kugel ohne hinsehen?

Zusammenfassung Sortieren

- Sortieren geht recht schnell, eine Million Elemente in 0.1 Sekunden auf Notebook
- Weltrekorde
 - Eine Billion Zahlen in drei Stunden
 - 50 Milliarden Zahlen für einen Penny

Suchbäume

2,7,3,9,4, Suche nach 9, 11,1,6, Suche nach 8

Zusammenfassung

- Binärsuche ist rasend schnell: 40 Vergleiche für Suche in einer Billion Elemente
- Sortieren ist billig: eine Million Elemente in 0.1 sec auf diesem Notebook
- Suchbäume erlauben Binärsuche auf dynamischen Daten