

*“Wer Ordnung hält,
ist nur zu faul zum Suchen...”*

**Wie schnell findet
Facebook einen von
2,5 Milliarden Benutzern?**

Klaus Kusche
Frühjahr 2019

Suchen? Computer rechnen doch!

Mehr rechnen oder mehr suchen???

- Wettervorhersage
- Google, Facebook, YouTube, Wikipedia, ...
- Telefon-Auskunft, Zug-Auskunft, ...
- Online-Shop
- Rechtschreibprüfung
- Robotersteuerung
- Bank, Finanzamt, Krankenkasse, ...

Suchen? Computer rechnen doch!

Mehr *rechnen* oder mehr *suchen*???

- **Wettervorhersage**
- Google, Facebook, YouTube, Wikipedia, ...
- Telefon-Auskunft, Zug-Auskunft, ...
- Online-Shop
- Rechtschreibprüfung
- **Robotersteuerung**
- Bank, Finanzamt, Krankenkasse, ...

Was ist daran schwer?

*“Einfach alle der Reihe nach
durchsuchen,
bis man den Richtigen hat!”*

Weil:

*Der Computer hat keinen “Überblick”,
er kann nicht “alle auf einmal” anschauen,
sondern immer nur zwei Sachen
miteinander vergleichen!*

==> Zeichen für Zeichen, Name für Name!

Was ist daran schwer?

*“Einfach alle der Reihe nach durchsuchen,
bis man den Richtigen hat!”*

Annahmen:

- 2,5 Milliarden Namen mit im Mittel 10 Zeichen
- Alle Namen haben gleichzeitig im RAM Platz
 - Braucht viele GB RAM!
 - Platte/SSD wäre viel langsamer & komplizierter
- Prozessor mit 4 Ghz Takt

Was ist daran schwer?

*“Einfach alle der Reihe nach durchsuchen,
bis man den Richtigen hat!”*

Ganz grob:

Vergleich braucht 10 Takte pro Zeichen

==> 100 Takte pro Namen (10 Zeichen)

==> 25 Nano-Sekunden pro Namen

(4 000 000 000 Takte/Sek, d.h. 1 Takt = 0,25 Nano-Sek)

==> 62,5 Sekunden für 2 500 000 000 Namen!!!

Muss man wirklich alle anschauen???

Bei “zufälliger Reihenfolge” der Namen:

- Namen ***vorhanden***:

Man findet den Namen im Mittel
nach der **Hälfte** aller Namen

=> *Immer noch zu langsam!*

- Namen **nicht vorhanden**:

Das kann man erst dann sagen,
wenn man **alle** Namen angeschaut hat!

Kann man “*cleverer*” suchen?

In einem “Saustall”:

Nein!

Auch ein Computer kann nicht “*zaubern*”...

Kann man “*cleverer*” suchen?

In einem “Saustall”:

Nein!

Auch ein Computer kann nicht “zaubern”...

Aber:

Ordnung hilft ja
angeblich beim Suchen!

Was heißt “*Ordnung*” ?

Ordnung = Sortierung

bei uns:

- ... nach Name *alphabetisch*

in anderen Fällen:

- ... nach Artikel*nummer*,
Versicherungs*nummer*,
Konto*nummer*, ...
- ... nach *Datum*

Was bringt Ordnung?

Für die “*dumme*” *Suche*
(einfach *der Reihe nach* durchschauen):

(Fast) nichts!

(nur der Fall “*nicht vorhanden*” wird doppelt so schnell)

Was bringt Ordnung?

Für die “*dumme*” *Suche*
(einfach *der Reihe nach* durchschauen):

(Fast) nichts!

(nur der Fall “*nicht vorhanden*” wird doppelt so schnell)

Aber wenn die Daten sortiert sind:

Jetzt kann man “cleverer” suchen!

Wie???

“Divide et impera”!

“Teile und herrsche!”

Der Trick heißt

“binäre” Suche

“binär” hat etwas mit ***“zwei”*** zu tun...

“Divide et impera”!

... also immer wieder

zweiteilen:

In der richtigen *Hälfte*, im richtigen *Viertel*,
im richtigen *Achtel* usw. weitersuchen,

bis

- der Name *gefunden* ist
- oder *nichts mehr übrig* ist
(dann gibt es den Namen nicht!)

“Divide et impera”!

Aber was ist die “richtige” Hälfte?

Den Namen ***in der Mitte***
des zu durchsuchenden Teils
mit dem gesuchten Namen vergleichen

- Mitte ist ***gleich*** gesuchtem Namen: ***Fertig!***
- Mitte ist ***kleiner*** als der gesuchte Name:
Gesuchter Name muss in ***rechter Hälfte*** sein!
- Mitte ist ***größer*** als der gesuchte Name:
Gesuchter Name muss in ***linker Hälfte*** sein!

Was bringt “Hirn” & “Ordnung”?

Mit jedem einzelnen Vergleich mehr kann man doppelt so viele + 1 Namen durchsuchen!

<i>... Vergleiche</i>	<i>reichen für ... Namen!</i>
1	1
2	3
3	7
4	15
5	31
6	63
10	1023
20	1048575
30	1073741823

Was bringt “Hirn” & “Ordnung”?

Für 2,5 Milliarden Facebook-Benutzer reichen

32 Vergleiche,

um den richtigen zu finden!

800 Nano-Sekunden statt 1 Minute,
daher über 75 Millionen Mal schneller!

Alles bestens?

Leider noch nicht:

So wie sich das Kinderzimmer
nicht von selbst aufräumt...

*... sortieren sich auch die Daten
nicht "von selbst"!*

Sortierte Reihe von Schülern

- 1.) Die Kinder kommen einzeln und in zufälliger Reihenfolge:

Sortierte Reihe von Schülern

- 1.) Die Kinder kommen einzeln und in zufälliger Reihenfolge:
 - Jeder sucht sich von hinten der Reihe nach den richtigen Platz
 - Alle größeren Namen rutschen 1 Platz nach hinten

“Sortieren durch Einfügen”

Sortierte Reihe von Schülern

2.) Die Kinder sind schon alle da,
aber stehen auf einen Haufen durcheinander:

Sortierte Reihe von Schülern

- 2.) Die Kinder sind schon alle da,
aber stehen auf einen Haufen durcheinander:
- Der Lehrer sucht der Reihe nach aus allen den ersten Namen, den zweiten Namen, usw.
 - Jeder Ausgesuchte stellt sich hinten in die Reihe,
keiner muss nachrutschen

“Sortieren durch Auswählen”

Sortierte Reihe von Schülern

3.) Die Kinder stehen schon
durcheinander in einer Reihe,
jeder sieht *nur seine beiden Nachbarn:*

Sortierte Reihe von Schülern

3.) Die Kinder stehen schon
durcheinander in einer Reihe:

- Der Lehrer geht immer wieder *von links nach rechts* durch die ganze Reihe
- Wenn zwei Nachbarn *“falsch”* stehen, wird *Platz getauscht*
- ... bis alles stimmt (ein Durchgang ohne Tauschen)

“Sortieren durch Vertauschen”

Klappt das auch im Computer?

Eigentlich *ja, aber*

alle 3 Sortierverfahren

brauchen für 2.500.000.000 Namen in etwa ...

Klappt das auch im Computer?

Eigentlich ja, aber

alle 3 Sortierverfahren

brauchen für 2.500.000.000 Namen in etwa ...

$$2.500.000.000 * 2.500.000.000 / 2$$

Vergleiche!!!

Das wären 78.125.000.000 Sekunden
oder rund 2.500 Jahre!

Hoffnung Nummer 1

Wenn man bei 2.500.000.000 Namen einen Namen mit 32 Vergleichen suchen kann, sollte man auch

*mit 32 Vergleichen
den richtigen Platz
für einen einzelnen Namen*

finden können!

Das wären 80.000.000.000 Vergleiche für 2.500.000.000 Namen (oder gut 30 Minuten...)

Hoffnung Nummer 2

Hilft

“teile und herrsche”

auch beim Sortieren?

==> *Wie könnte man teilen???*

Sortieren durch Teilen...

- Einen Namen zufällig auswählen
(hoffentlich erwischt man einen “mittelgroßen” Namen, sonst werden die Hälften recht ungleich!), ...
- ... alle kleineren Namen links davon und alle größeren Namen rechts davon hinstellen
(irgendwie, unsortiert), ...
- ... und jede Hälfte wieder genauso weiter aufteilen...
- ... bis jede “Gruppe” nur mehr einen Namen groß ist

Jetzt wären wir berühmt ...

... wenn wir das
vor über 50 Jahren erfunden hätten:

“Quicksort”

(Tony Hoare, 1960,
wurde Professor an der Universität Oxford
und Chef eines Microsoft-Forschungslabors)

Heute das häufigste Sortierverfahren

... und wirklich so schnell wie erhofft!

Die Wissenschaft hat jedenfalls bewiesen:

Es ist unmöglich, deutlich schneller zu sortieren!

Aber jetzt ist alles bestens...

... oder doch nicht???

Aber jetzt ist alles bestens...

... oder doch nicht???

*Was ist,
wenn sich die Daten oft ändern?*

Jedesmal, wenn einer dazukommt / wegfällt:

Aber jetzt ist alles bestens...

... oder doch nicht???

*Was ist,
wenn sich die Daten oft ändern?*

Jedesmal, wenn einer dazukommt / wegfällt:

Alle dahinter um eins verschieben!

Bei Facebook:

Jedesmal ~1.250 Millionen Benutzer umkopieren!

*Viel zu aufwändig,
dauert auch Minuten!!!*

“Brief ans Christkind”

- Ich möchte neue Namen speichern können, wo gerade Platz ist, ganz durcheinander!
- Ich möchte Namen einfach weglöschen können, auch wenn dadurch Löcher in meiner Liste entstehen.
- Ich möchte aber trotzdem binär suchen können!
- Und ich will auch eine sortierte Liste aller Namen ausdrucken können!
- Und das alles ohne jemals alle Daten zu sortieren oder zu verschieben!

“Binär suchen” braucht ...

... für jeden Namen

... nur den mittleren Namen
des Teils links davon

... und den mittleren Namen
des Teils rechts davon!

Trick: “Hausnummern”

- 1.) Jeder Speicherplatz bekommt eine ***eindeutige “Hausnummer”***
- 2.) Bei jedem Namen speichern wir zwei Hausnummern anderer Namen:

Wo “wohnt” der mittlere Name seines linken Teils und seines rechten Teils?

- 3.) Wenn es links oder rechts von einem Namen keinen Teil (keine Namen) mehr gibt, tragen wir **“0”** als Hausnummer für den leeren Teil ein.

Informatiker sind komische Leute...

- So eine Anordnung von Daten heißt

Baum

genauer “binärer Suchbaum”.

- Für jeden Namen gilt:

Alle kleineren Namen “hängen” links an ihm dran,
alle größeren rechts.

- Gezeichnet werden Bäume meist mit der “Wurzel” (= mittlerer Namen) oben, Wuchsrichtung nach unten!

... und sie klauen Ideen gerne

Bäume gibt es in vielen Varianten,

und sie stecken
in fast jedem Programm,

das große Datenmengen

speichert und durchsucht!

(z.B. auch in allen Datenbanken)

“The end”

Fragen?