

Wo liegt eigentlich dieses Internet?



Daten,Daten,Daten

Ich

- ... weiß wie eine Suchmaschine funktioniert (Crawling,Indexing)
- ... kenne verschiedenen Speicherarten
- ... kann lokale Daten und Daten im Internet unterscheiden
- ... weiß was ein Betriebssystem ist und was Anwendersoftware ist
- ... kenne Grundbegriffe ...

Datenzentrum



Unsere Daten liegen also strukturiert in Datenbanken. Aber wo liegen diese Datenbanken mit all diesen Daten? Wenn wir bei google etwas suchen, gibt es meist Millionen von Treffern. Wenn wir die alle in eine Datenbank legen würden, wird das sicher sehr lange dauern, bis wir da drin etwas finden.Und auf einem einzelnen Computer würde der Speicherplatz wohl nicht ausreichen. Also hat man sich so etwas wie ein *Datenzentrum* ausgedacht. Such den Begriff doch mal, Du wirst überrascht sein. Viele Firmen die so viele Daten haben und/oder viel Rechenpower benötigen, besitzen eigene Rechen-/Datenzentren. Facebook zum Beispiel oder Google. Google muss ja das ganze Internet speichern können um Suchergebnisse anzuzeigen.

Schau Dir an wie es im **Facebook-Rechenzentrum** in Europa aussieht
Ein weiteres großes Rechenzentrum hat natürlich **Google**

S..S..Server...s..s..schrank?



In so einem Datenzentrum stehen also solche Schränke mit ganz vielen Computern drin. Laufen wir doch mal durch ein solches Zentrum von Google.

Inside Google

Da ist ja alles voll mit Schränken. Halte doch mal vor so einem Schrank an und schau was da so drin steckt.

Hintergrundinformationen



In so einem Schrank sind anscheinend viele Computer untergebracht und miteinander vernetzt. Warum haben diese Computer keinen Bildschirm und keine Tastatur oder andere Geräte?



Das geht schon, aber es ist nicht notwendig. Wenn man unbedingt an einen dieser Computer heran muss, kann man auch einen z.B. Laptop anschließen. Die Computer werden sonst aus der Ferne über Programme gesteuert. Und

auf solchen Computern lagert z.B. Google dann das Internet in Datenbanken.



Im Rausch der Geschwindigkeit



Jetzt wissen wir also wie die erste Zahl zustande kommt - die Anzahl Suchergebnisse. Aber wie kommt die zweite Zahl zu Stande, die Zeit. 0.53 Sekunden? Falls Du Dir schon [Datenstrukturen](#) angesehen hast, dann weißt Du, dass wir so große Datenmengen sehr schnell durchsuchen können, wenn wir sie strukturieren z.B. in Bäumen. Aber wer durchsucht das denn? Das macht ein Computer. Der muss demnach "schnell" sein. Was heißt das? Hast Du eine Ahnung was einen schnellen



Computer ausmacht? Was braucht unser Computer um die Suchmaschine möglichst schnell zu machen?

Speicherkapazität



Wir brauchen als erstes einmal sehr viel Speicherplatz. Weißt Du, wieviel Dein Smartphone oder Dein Computer an Speicherplatz besitzen?

Eine **Festplatte** hat heutzutage



ungefähr einen Terabyte. Das sind 1024 Gigabyte. Und das sind 1.048.576 Megabyte.... und Du hast mich sicher verstanden. Das sind eine Menge Bits und Bytes. Damit die Festplatte die Daten lesen kann, bewegt sie diesen Arm auf der Scheibe an den Punkt, wo die Daten liegen. Du kannst Dir sicher vorstellen, dass die Bewegung des Armes ein wenig Zeit braucht.

Wenn nun Millionen von Menschen bei google eine Suche starten, könnte es mitunter sehr lange dauern, bis diese Arme zu den richtigen Stellen finden.

Geht das nicht schneller? Geht es. Mit Hauptspeicher (RAM).



In den Hauptspeicher werden die Daten gelegt, die öfter gebraucht werden. Damit sie nicht ständig von der Festplatte gelesen werden müssen. Nachteil: Machst Du den Strom aus, entleert sich auch der Hauptspeicher.

Rechenkapazität



Okay. Schneller Speicher. Check. Was brauchen wir noch? Wir brauchen noch so etwas wie das Gehirn eines Computers - einen Prozessor



Da drin laufen alle Berechnungen ab, damit der Computer und seine Schaltkreise wissen, was zu tun ist. Die typische Architektur wird als die von Neumann-Architektur beschrieben. Das ist Mathematiker, der sich überlegt hat, wie so ein Computergehirn funktionieren müsste.

Zu seiner Zeit hatte ein Computer meist nur ein Programm, dass er ausführen konnte. Das war durch seine Hardware fest mit ihm verbunden oder er musste es von einer Lochkarte lesen. Wenn Du genauer wissen möchtest wie das funktioniert, schau in das folgende Video hinein.



<https://youtube.com/watch?v=YgdbUHUETBA>

Mit der von Neumann-Architektur war es zum ersten Mal möglich mehrere Programme ausführen zu lassen und Programme zu ändern. Das bist Du heute gewohnt von Deinen Geräten. Du hörst Musik, spielst Spiele und surfst im Internet. Für all das brauchst Du keine 3 Geräte, Du kannst es mit einem erledigen. Aber wer sagt denn, welches Programm dran ist, wenn es nur einen Prozessor gibt?

Was läuft auf der Kiste?



Auf Deinem Smartphone läuft vielleicht gerade Musik, Du chattest mit einem Freund / einer Freundin und da kommt auch noch ein Anruf rein. Wieso funktioniert das? Du hast gesehen, dass durch die von Neumann-Architektur möglich wird, mehrere Programme über einen Prozessor auszuführen.

Darum kümmert sich das **Betriebssystem**.

Auf dem Bild siehst Du einen Task-Manager. so etwas kennst Du vielleicht. Hier

Name	15% CPU	39% Arbeitsspeicher	9% Datenträger	0% Netzwerk
Apps (7)				
> Google Chrome (32 Bit)	0,4%	386,9 MB	0 MB/s	0,1 MBit/s
> NoMachine Player Process (32 ...)	1,8%	79,2 MB	0 MB/s	0,1 MBit/s
> paint.net (3)	0%	74,3 MB	0 MB/s	0 MBit/s
> Skype (32 Bit) (19)	3,8%	106,6 MB	0,1 MB/s	0 MBit/s
> Snipping Tool	0%	2,5 MB	0 MB/s	0 MBit/s
> Task Manager	2,5%	12,2 MB	0 MB/s	0 MBit/s
> Windows-Explorer (2)	2,0%	89,0 MB	0 MB/s	0 MBit/s
Hintergrundprozesse (59)				
Alps Pointing-device Driver	0%	0,5 MB	0 MB/s	0 MBit/s
Alps Pointing-device Driver	0,6%	1,7 MB	0 MB/s	0 MBit/s
Alps Pointing-device Driver for ...	0,6%	0,7 MB	0 MB/s	0 MBit/s
ApMsgFwd	0,8%	0,5 MB	0 MB/s	0 MBit/s
Application Frame Host	0%	5,9 MB	0 MB/s	0 MBit/s

siehst Du welche Programme gerade laufen. Wieviel Hauptspeicher sie belegen und wieviel Leistung sie vom Prozessor beanspruchen.

Das Betriebssystem regelt also u.a. welcher Prozess wie lange den Prozessor beanspruchen darf. Es wechselt also so schnell die Prozesse, dass es uns vorkommt, alles würde zu gleich laufen. Sicher hast Du schon bemerkt, dass die Musik manchmal aussetzt, wenn Du gerade eine andere App lädst. Weil dann das Laden dieser App kurz etwas mehr Zeit beim Prozessor benötigt hat.

Wir haben also das Betriebssystem das unseren Computer verwaltet und wir haben

Anwendersoftware (Apps,Programme). Was kennst Du eigentlich für Betriebssysteme?

Übrigens wundere Dich nicht wegen **Programm** oder **Prozess**. Ein Programm wird Prozess genannt, wenn es gerade ausgeführt wird.

In den Rechenzentren in der Serverschränken laufen also Computer mit Betriebssystemen und Programmen. Aber keine Videospiele oder Musik, sondern Programme wie zum Beispiel die Datenbanken. Oder Programme die Daten aus den Datenbanken herausholen und verarbeiten.

Datenübertragungsrate



So. Schneller Speicher (Hauptspeicher), Prozessor, Betriebssystem und Programme. Haben wir alles. Was braucht es noch? Nun ja, die Computer müssen mit dem Internet verbunden sein.

Ist Dir schon einmal aufgefallen, wenn Du google.com eingibst, dass Du bei google.ch landest? Wenn Du das in Deutschland versuchst, landest Du auf google.de. D.h. wenn Du Dich im Internet bewegst, gibt sich Dein Computer bzw. Dein Router zu

erkennen mit einer IP-Adresse.

Du wirst also auf ein Rechenzentrum umgeleitet, was näher zu Dir ist, als das in Amerika (google.com). Sonst würden die Daten zu Dir einen größeren Weg zurücklegen müssen, als aus Europa.

Du kennst das vielleicht aus Spielen. Wenn Du auf einem Server spielst, der z.B. in Amerika oder Russland liegt, dann **lagst** Du. D.h. es gibt eine spürbare Verzögerung zwischen Dir und dem Server.



https://youtube.com/watch?v=_fNp37zFn9Q

Das Video veranschaulicht den Effekt mit einem lustigen Experiment. Es ist also wichtig, dass die Datenübertragungsrate hoch ist, und die Server in Deiner Nähe stehen. Darum haben viele große Unternehmen ihre Rechenzentren überall auf der Welt verteilt, damit ihre Anwendungen von überall schnell erreichbar sind.

Und wenn ein Teil des Datenzentrums ausfällt?



Dann würde das bedeuten, wenn ein paar dieser Server ausfallen, ist ein Teil einer Anwendung, zum Beispiel Facebook nicht mehr erreichbar. Oder wenn bei Google etwas ausfällt, dass dann nicht mehr alle Suchergebnisse erscheinen.

Das ist prinzipiell richtig. Deswegen gibt es Strategien sich davor zu schützen. Zum Beispiel **Backups**. Dabei geht es um **Redundanz**. Das heisst einfach, die Daten liegen mehrfach auf verschiedenen Servern. So gibt es eine **Ausfallsicherheit** und die Daten können im Falle eines Ausfalls woanders abgerufen werden.

Das gleiche machst Du vielleicht mit Deinen Bildern. Die sicherst Du auch an mehreren Stellen oder direkt in der Cloud. Damit im Backup immer alle z.B. Bilder vorhanden sind, musst Du aber oft **synchronisieren**. Das heisst Dein Backup auf den aktuellen Stand bringen.

Manchmal kann es aber sein, dass Du gern einen anderen (meist alten) Stand der Daten brauchst. Dafür gibt es die **Versionierung**. So werden zu einem Datenbestand mehrere **Versio**nen angelegt. Zum Beispiel könnte jedes neue Backup eine neue Version sein. So kannst Du beliebig zwischen Versionen hin und her wechseln.



Ein prominentes Beispiel ist die Wikipedia. Du kannst dort zu jedem Artikel die Versionsgeschichte genauer betrachten. Du siehst, wer wann welche Änderungen durchgeführt hat. Da jeder in der Wikipedia mitschreiben darf, können so unerwünschte Änderungen rückgängig gemacht werden, indem ein vorheriger Stand des Artikels wiederhergestellt wird aus einem (Achtung, ein Satz mit allen Begriffen) Backup was zuvor synchronisiert wurde und somit eine ältere Version besitzt.

Wo kommen die Daten der Suchmaschine her?



Wenn ich ein Foto ins Internet stelle, warum findet die Suchmaschine das?
Eine (fast) letzte große Frage Wie kommen eigentlich die ganzen Daten zusammen?
Wer sammelt die damit sie bei google landen können?
Du hast es sicher schon geahnt - das machen Programme. Sogenannte Crawler. Mit Soekia steht Dir eine Webanwendung zur Verfügung, mit der Du das Prinzip von Suchmaschinen versuchen kannst nachzuvollziehen.

Soekia 2.0

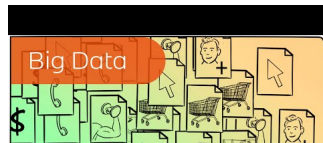
Mit Soekia kannst Du hinter die Kulissen von Suchmaschinen schauen und damit deren Grundprinzipien kennenlernen.

▶ SOEKIA STARTEN

Big Data



Warum weiß die Suchmaschine, dass ich neue Sportschuhe brauche?
Die Frage ist sehr interessant, aber auch nicht einfach zu beantworten. Vielleicht hilft Dir das folgende Video weiter. Du hast es eventuell auch schon in einem anderen Arbeitsblatt gesehen.



https://youtube.com/watch?v=uH813u7_b0s