

03.03.13 | Technik

## Wie gewaltig Big Data unser Leben verändert

Unfassbar schnelle Computer und neuartige Software können mit riesigen Datenmengen umgehen. Das verändert nicht nur die Naturwissenschaften, sondern auch unser Leben und unsere Art zu denken. Von *Thomas Jüngling*

Ganz wundersame Dinge gehen vor sich: Das US-Agrarministerium hat fast die komplette Online-Kommunikation zwischen den Landwirten in den USA gesammelt und ausgewertet. Aus den Beiträgen der Bauern – einem riesigen Datenberg – wollen sie herausfinden, wie sich Pflanzen optimal für einen bestimmten Bedarf züchten lassen. Hier war etwas am Werk, das auf der nächste Woche startenden Cebit das ganz große Thema ist: Big Data.

Mit neuen Analysemethoden und innovativer Technik ist es möglich, für das menschliche Gehirn nicht mehr fassbare Datenmassen aus unterschiedlichsten Quellen zu speichern, in ihnen nach Mustern zu suchen und daraus Schlüsse zu ziehen. Experten aus allen Bereichen sind sich sicher, dass dies unser Leben fundamental verändern wird, unser Konsumverhalten, unseren Umgang mit Krankheiten, unsere Art zu denken.

"Big Data revolutioniert außerdem gerade die Naturwissenschaften", sagt Peter Sanders, Informatik-Professor am Karlsruher Institut für Technologie. So sei zum Beispiel die Entdeckung des Higgs-Teilchen am Cern nicht ohne Big Data möglich gewesen.

### So viele Daten wie nie zuvor

Es kommen massenhaft Daten in die Welt. Sie stammen unter anderem von zahlreichen Sensoren, die in Maschinen stecken. Sie erfassen alles Messbare, von Bewegungen über Temperaturen bis zu chemischen Veränderungen in der Luft. Dazu kommen Informationen von Barcodes und Funketiketten. Da auch Menschen viel über sich im Netz preisgeben, kommen hier nochmals so viele Daten wie niemals zuvor zusammen.

Das katapultiert die Datenmenge in schwindelerregende Zahlen-Dimensionen. 2011 belief sich der Datenbestand auf 1,8 Zettabyte – dahinter steckt eine Zahl mit 21 Nullen. 90 Prozent dieses Datenmonstrums wurden in den vergangenen zwei Jahren erzeugt.

Oder anders ausgedrückt: Allein im vergangenen Jahr haben Menschen und Maschinen so viele – verfügbare und zugängliche – Daten produziert wie in der gesamten Menschheitsgeschichte zuvor. Sie würden nur knapp auf die Speicher von 60 Milliarden Tablet-Computern passen. Täglich kommen 2,5 Exabytes hinzu, und das Volumen neu anfallender Daten verdoppelt sich alle zwei Jahre.

### Wenige strukturierte Daten

Nur 15 Prozent dieser Daten sind nach einer Berechnung von TNS Infratest strukturiert und sauber, lassen sich also mit herkömmlicher Software in Tabellen einsortieren. Der "schmutzige" Rest ist eine wilde Mischung aus E-Mails und Präsentationen, aus Facebook-Einträgen und Stellengesuchen, aus Laufzeiten eines Trainings und Geboten bei Ebay, aus

Anzeigen bei Singlebörsen und Sucheingaben bei Google. Ein weiteres Merkmal der aktuellen Entwicklung: Die Daten stehen in einem atemberaubenden Tempo zur Verfügung, die Datenströme haben sich extrem beschleunigt.

Darauf haben sich IT-Experten weltweit mittlerweile eingestellt. Eine technische Lösung dafür lautet paralleles Arbeiten. Da die Prozessoren in Rechnern kaum noch schneller werden, lassen sie einfach mehrere von ihnen gleichzeitig an einer Aufgabe arbeiten – eben auch bei der Analyse riesiger Datenmengen.

Eine weitere Innovation heißt "In-Memory". Dabei wird der Arbeitsspeicher eines Computers als Datenspeicher eingesetzt. SAP und IBM zum Beispiel schließen mehrere Rechnereinheiten mit jeweils zehn Rechenkernen sowie die Server untereinander zusammen – und so kommt der Arbeitsspeicher auf satte 250 Terabyte. Zum Vergleich: In einem iPad werkelt ein Speicher mit einem Gigabyte. Er ist also 250.000-mal kleiner.

## **Neue statistische Verfahren**

Mit diesen Rechenmaschinen ist es also möglich, 250 Terabyte an Daten auf einmal zu verarbeiten. Das wäre so, als würde man während eines Wimpernschlags einige Millionen Bücher lesen. Nicht nur leistungsfähige Maschinen, auch neue statistische Verfahren sowie neue Software sind für Big Data notwendig. Die klassischen Statistik-Algorithmen und Programme wie Excel kommen mit den Datenmassen nicht zurecht. Big-Data-Programme verteilen die Aufgaben auf ein Netzwerk aus Computern. Danach führt eine Software die Ergebnisse zusammen, um nach Mustern zu suchen oder Daten auf bislang unbekannte Weise miteinander zu verknüpfen.

Im Gegensatz zu früher kann die Software, basierend auf semantischer Technologie, noch mehr: Sie unterscheidet gleichlautende Begriffe mit unterschiedlicher Bedeutung, sie verknüpft Daten miteinander, die scheinbar nichts miteinander zu tun haben – und finden so verborgene Muster.

IBM-Forscher haben zum Beispiel entdeckt, dass Babys mit regelmäßigem Herzschlag später mit höherer Wahrscheinlichkeit erkranken als Babys mit unregelmäßigem Herzschlag. Nicht nur der Befund mag überraschen, auch der Weg dahin ist ungewöhnlich: Es waren keine Ärzte, die ihre Erkenntnisse zusammengetragen haben. Eine Software hat den Zusammenhang entdeckt.

## **Wartungs-Informationen während des Flugs**

Anwenden lassen sich die neuen Methoden in fast allen Bereichen. Sensoren einer Boeing-Turbine zum Beispiel liefern jede Stunde 20 Terabyte, um schon während des Flugs Informationen für Wartungstechniker zu sammeln, die später am Boden Hand anlegen – und bereits vor der Landung die passenden Ersatzteile organisiert haben.

Windrad-Hersteller Vestas kann mit einer Big-Data-Software an einem Standort berechnen, wie hoch der Ertrag einer Windkraftanlage dort sein würde, wie viel Energie sie erzeugen könnte, wie oft Reparaturen fällig wären. Insgesamt berücksichtigt das Programm 160 Faktoren, unter anderem Temperatur, Feuchtigkeit, Windrichtung, Waldkarten, Gezeiten und Satellitenbilder.

Big Data spielt in zwei Bereichen eine besonders große Rolle. Zum einen in der Wirtschaft. Unternehmen können an Kunden- und Konsumdaten ihre Materialbestellungen und ihr Produktangebot ausrichten. Außerdem ist der analytische Umgang mit Datenmassen in der medizinischen Forschung populär geworden.

## **Tweets kurbeln Medikamentenproduktion an**

Wenn zum Beispiel viele Menschen einer Region bei Google nach gewissen Medikamenten suchen und sie in ihren Benachrichtigungen bei Twitter über Kopfschmerzen

(Link: <http://www.welt.de/themen/kopfschmerzen/> ) klagen, dann könnte eine Grippewelle anrollen.

Hausärzte wären vorgewarnt, Pharmakonzerne könnten die Produktion entsprechender Medikamente ankurbeln.

Siemens-Techniker haben eine Software entwickelt, die in Dokumenten unterschiedlichster Art nach inhaltlichen Verweisen auf Gene sucht und so einen Zusammenhang zwischen einem bestimmten Gen und der Erkrankung an Alzheimer (Link: <http://www.welt.de/themen/alzheimer/> ) - gefunden.

Personalisierte Medizin soll mit Big Data möglich sein, unter anderem eine individuelle – und somit effektive – Krebstherapie. Manche Wissenschaftler schwärmen, damit könnten Menschen 200 Jahre und älter werden. "Mit Big Data wird bald eine Sequenzierung des Genoms nicht mehr kosten als eine Computertomografie. Das wäre unschätzbar wertvoll für eine verbesserte Diagnose und der Personalisierung von Medikamenten-Therapien", sagt Sanders.

## Schattenseite beim Gen-Sequenzieren

Doch das Beispiel zeigt auch die Schattenseite von Big Data. Die für fast jeden erschwingliche Möglichkeit zur Genom-Sequenzierung würde den Druck erhöhen, diese tatsächlich vornehmen zu lassen: Wer will sich schon nachsagen lassen, dies nicht bei seinem Kind veranlasst zu haben, und wie will er das vor seiner Krankenversicherung rechtfertigen? Außerdem sind mit Big Data massive Datenschutzprobleme verbunden.

Mitglieder sozialer Netzwerke zum Beispiel geben Informationen schließlich nicht preis, damit Handelsketten oder Krankenversicherungen sie nutzen. Das ist auch rechtlich heikel: "Zunächst werden die Daten aus ihrem ursprünglichen Kontext gelöst, also der Grundsatz der Zweckbestimmung verletzt, wonach personenbezogene Daten nur zu den Zwecken verwendet werden dürfen, zu denen sie erhoben worden sind", führt der Bundesdatenschutzbeauftragte Peter Schaar aus.

Big Data kann auch die Arbeit von Wissenschaftlern verändern. Für manche ist der Fall klar: Je mehr Daten, desto höher die wissenschaftliche Erkenntnis. Viel mehr als Daten brauche man nicht. Chris Anderson, Ex-Chefredakteur des Magazins "Wired", hat sogar schon das "Ende der Theorie" ausgerufen.

## Widerspruch der Theoretiker

Ein Ende der Theorie sei nicht zu erkennen, widerspricht Klaus Mainzer, Wissenschaftstheoretiker an der TU München, und verdeutlicht dies anhand der Molekularbiologie: "Theorien komprimieren viele Daten in kürzeren Formeln, um Voraussagen zu machen. Die Vielfalt der komplexen Wechselwirkungen zwischen Zellen zu erfassen, fällt allerdings mit den exakten naturwissenschaftlichen Methoden sicherlich schwer. Doch wer nur nach statistischen, korrelativen Zusammenhängen sucht, stochert oft mit der Stange im Nebel. Es führt kein Weg daran vorbei, Gesetzmäßigkeiten zu finden und sie zu Theorien zu verknüpfen – auch, um zum Beispiel verlässliche Therapien gegen Krebserkrankungen entwickeln zu können." Dazu reiche die Erkenntnis nicht aus, *dass* bestimmte Wechselwirkungen unter definierten Bedingungen auftauchen, sondern Forscher müssen auch ergründen, *warum* dies der Fall sei.

"Würden wir nur den Daten hinterherlaufen, wäre die Wissenschaft am Ende. Allerdings können Algorithmen Muster erkennen, die Forschern helfen, neue Hypothesen zu bilden. Computer sind sehr nützlich, um Vorschläge zu machen", sagt Mainzer. Bei aller Leistungsfähigkeit können Algorithmen jedoch nicht kritisch auf einen Befund schauen, den

ihnen die Datenanalyse geliefert hat. Forscher sind in der Lage, Erkenntnisse skeptisch zu beurteilen, Software kann dies nicht. Zumindest noch nicht: Programmierer arbeiten auch daran.

---

© Axel Springer SE 2014. Alle Rechte v