

Wo Riesendaten der Menschheit helfen

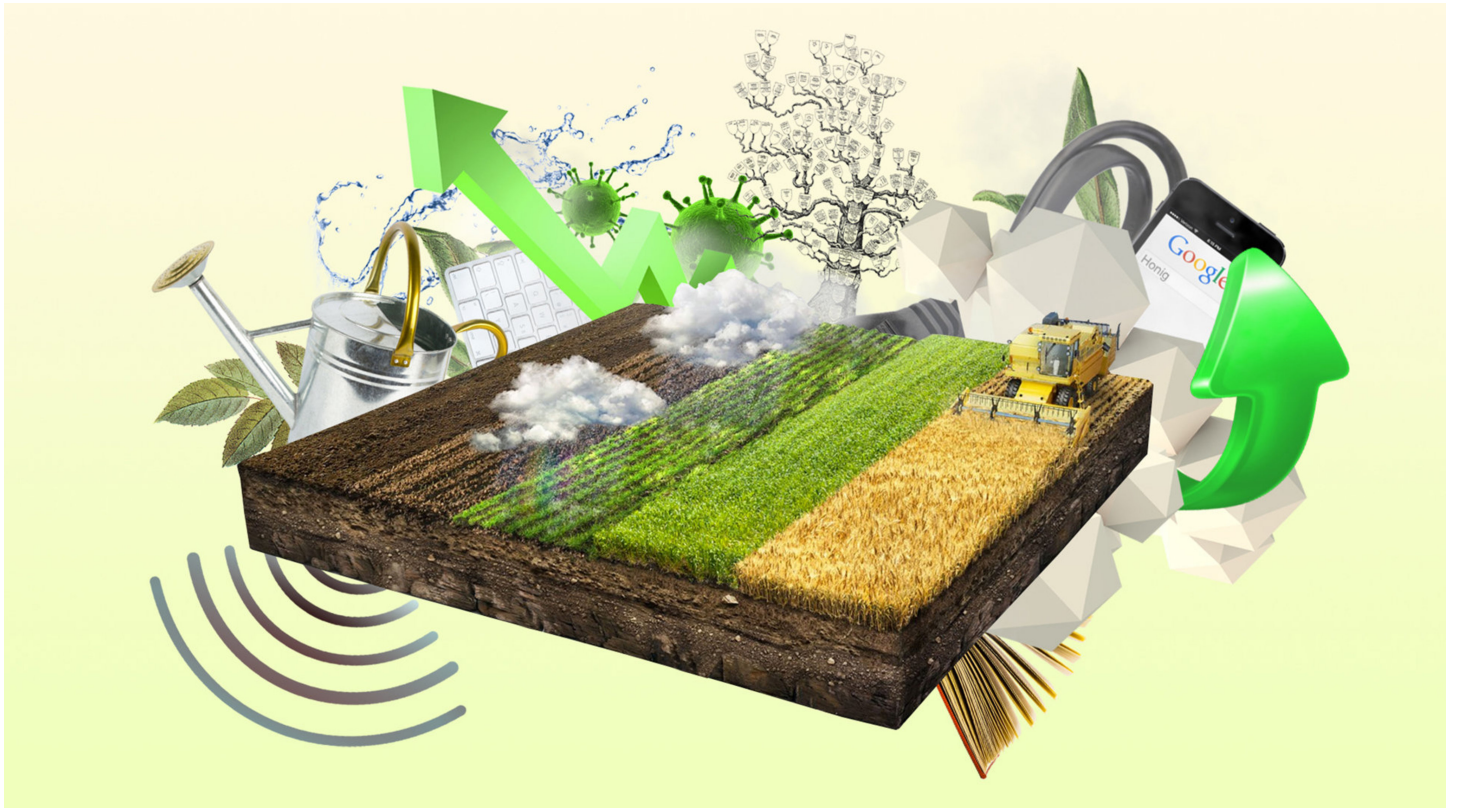
Bei Big Data denkt man an mächtige Internetkonzerne und den Verlust von Privatsphäre. Aber im großen Datensammeln liegen auch Chancen. Fünf Beispiele

Bernd Kramer

31.10.2018

Daten

5 Min.



Mit Sensoren auf den Feldern die Ernte verbessern

Wie kann eine wachsende Weltbevölkerung ernährt werden, wenn die Ackerflächen auf der Erde begrenzt sind? Landwirte versuchen ihre Erträge zu erhöhen, indem sie sehr umfangreiche Daten auswerten. Mit speziellen Sensoren in den Landmaschinen lässt sich zum Beispiel punktgenau die Bodenqualität erfassen. Entsprechend können Felder ganz gezielt gedüngt werden. Präzisionslandwirtschaft nennt sich diese Form des datengestützten Anbaus. Eine Studie des Europäischen Parlaments aus dem Jahr 2016 rechnete vor, wie die Landwirtschaft durch Big Data effizienter und umweltfreundlicher werden könnte:



Smart Farming schön und gut – persönliche Daten sollte man aber unbedingt schützen. Wie, erfährst du hier

Durch zielgenauere Düngung zum Beispiel ließen sich Stickstoffrückstände im Boden um 30 bis 50 Prozent reduzieren. Mehr als 80 Prozent der Unkrautbekämpfungsmittel ließen sich einsparen, wenn die Felder nicht breitflächig, sondern punktuell dort besprüht würden, wo es am nötigsten ist. Allerdings sollten die Daten nicht allein bei einem Großkonzern landen, der beispielsweise auch das Düngemittel verkauft. Mit dem Wissen, wann der beste Zeitpunkt zum Düngen ist, könnte das Unternehmen sonst nämlich die Preise gezielt nach oben schrauben, befürchten Kritiker. Die Innovationsinitiative Landwirtschaft 4.0 fordert deswegen eine öffentliche Förderung der vernetzten Landwirtschaft, um eine Abhängigkeit von Großunternehmen zu verhindern.

Mit Suchanfragen Krankheitsausbrüche vorhersagen

Eines der prominentesten Big-Data-Projekte in der Medizin hat sich zunächst einmal als Flop erwiesen: Google trat vor rund zehn Jahren mit einer Art Frühwarnsystem für Grippeepidemien an, das auf der Auswertung von Millionen Suchbegriffen US-amerikanischer Internetnutzer basierte. Die Annahme dahinter: Oft googeln Menschen ihre Symptome oder potenzielle Hausrezepte, bevor sie einen Arzt aufsuchen. Und damit auch, bevor die Gesundheitsämter feststellen können, dass sich Krankheitsfälle häufen und sie mit einer Impfkampagne reagieren sollten. Google Flu Trends, wie der Dienst hieß, wollte so fast in Echtzeit den Ausbruch ansteckender Krankheiten registrieren. Allein: Es klappte nicht so recht. Die Schweinegrippe-Pandemie 2009 übersah Google völlig, in anderen Fällen schlug der Dienst Alarm, obwohl keine Grippe im Anmarsch war, oder lag mit der Prognose weit daneben.

Die Geschichte zeigt ein grundsätzliches Problem vieler Big-Data-Ansätze: Die schier unermesslichen Datenmengen allein gewährleisten noch keine gute Vorhersage, man muss auch wissen, wie man in der Masse die wirklich aussagekräftigen Daten findet. Inzwischen hat Google den Dienst eingestellt. Dafür haben Statistiker der Universität Harvard die Idee aufgegriffen und verbessert: Google Flu Trends hatte zu Beginn für einen Zeitraum von fünf Jahren die meistgenutzten Suchbegriffe mit tatsächlichen Krankheitsdaten verglichen. Heraus kamen 45 Begriffe, die als Indikator für Grippe galten. Nur: Was in einem Jahr für eine Prognose taugt, ist im nächsten vielleicht schon wieder hinfällig. Die

Große Datenmengen bedeuten nicht automatisch große Erkenntnisse. Man muss sie erstmal richtig auswerten

Harvard-Wissenschaftler entwickelten deswegen ein dynamisches System.

Sobald sie neue Daten vom Zentrum für Krankheitskontrolle und Prävention bekommen, passen sie die Prognose an und berücksichtigen außerdem stärker, dass Menschen ihr Suchverhalten ändern: Die Forscher griffen auf die Suchanfrage-Informationen zurück, die Google öffentlich bereitstellt. Obwohl sie damit zwar weniger detaillierte Daten als Google Flu Trends zur Verfügung hatten, galt ihre Prognose als treffsicherer.

Mit digitalisierten Amtsunterlagen Ahnenforschung betreiben

Auch Historiker versuchen, immer größere Textmengen automatisch auszuwerten. Ein Beispiel dafür ist Venice Time Machine, ein Projekt, bei dem seit 2012 rund 190.000 Dokumente aus Venedigs Archiven eingescannt wurden. Wenn das Projekt abgeschlossen ist, kann jeder Internetnutzer in den Quellen und damit 1.000 Jahren Geschichte stöbern. Selbstlernende Algorithmen sollen in den Amtsunterlagen, Urkunden oder Briefen die Namen der Venezianer und ihre Verbindungen zueinander aufspüren. Könnten Computer eine vergangene Gesellschaft auferstehen lassen? Historiker träumen jedenfalls schon von einer „European Time Machine“ nach venezianischem Vorbild, die Rückschlüsse auf Sprachevolution, Krankheitsverbreitung und Entwicklungen in der Kunst oder Architektur zulässt.

Mit Handydaten Hungersnöte frühzeitig erkennen

Im Jahr 2009 gründeten die Vereinten Nationen Global Pulse, ein Institut, das sich mit der Analyse großer Datenmengen in der Entwicklungszusammenarbeit beschäftigt. Mittlerweile gibt es einige vielversprechende Ansätze. In einem Projekt etwa gingen die Forscher der Frage nach, ob sich aus Informationen über die Handynutzung frühzeitig ablesen lässt, wo Hungerskatastrophen drohen.

Wem das Geld für Essen fehlt, der läßt auch kein Guthaben auf sein Handy. Klingt logisch,

In dem ostafrikanischen Land, wo die Wissenschaftler ihre Untersuchung durchführten (welches genau, wurde nicht genannt), sind Mobiltelefone allgegenwärtig. Also ließen sich die Forscher von den Mobilfunkbetreibern anonymisierte Nutzerinformationen geben und verglichen sie mit den Ergebnissen einer Umfrage zum Lebensmittelkonsum, die Mitarbeiter

hilft NGOs aber dabei, schneller einzugreifen

des Welternährungsprogramms zur gleichen Zeit durchführten. Das Ergebnis: Viele der Menschen, die sich wichtige Grundnahrungsmittel nicht mehr leisten konnten, buchten auch besonders selten neues Gesprächsguthaben auf ihr Handy. Die Mobilfunkdaten können damit einen ersten Hinweis geben, wenn sich die Versorgungslage zuspitzt – sodass Organisationen wie das Welternährungsprogramm schneller als bisher mit Hilfsmaßnahmen beginnen können.

In einer anderen Fallstudie zeigte Global Pulse, dass sich in Indonesien, wo die Bevölkerung besonders twitterfreudig ist, aus Tweets praktisch in Echtzeit die Entwicklung der Lebensmittelpreise ablesen lässt – auch das könnte eine Warnung bei potenziellen Nahrungsmittelengpässen sein. Die Datenanalyse hatte allerdings eine Schwäche: Steigende Preise ließen sich aus den Tweets verlässlicher ableiten als fallende. Sobald das Essen teurer wird, setzen Twitter-Nutzer wütende oder besorgte Posts ab. Wird es billiger, nehmen sie das eher schweigend zur Kenntnis.

Mit digitalisierten Texten Muster in der Weltliteratur erkennen

Big Data – dahinter steht die Idee, mit immer leistungsfähigerer Software immer größere Datenmengen zu analysieren. Kein Wunder also, dass dieser Ansatz vor allem Vertreter von Disziplinen begeistert, die seit jeher mit statistischen Methoden arbeiten, wie etwa Naturwissenschaftler und Sozialforscher. Doch auch Geisteswissenschaftler entdecken die Datenanalyse – zumal ein wachsender Teil des Schrifttums der Menschheit inzwischen digital vorliegt. Forscher des Stanford Literary Lab beispielsweise haben 27.266 amerikanische Romane daraufhin analysiert, an welcher Stelle das Wort „Tod“ im Text mit der größten

Wahrscheinlichkeit auftritt. Das zugegebenermaßen wenig überraschende Ergebnis: Gegen Ende der Geschichte ist sehr viel häufiger vom Sterben die Rede als am Anfang – weil Romane oft mit dem Tod des Helden enden.

Aus einem ähnlichen Grund ballen sich auch Begriffe wie „Waffe“ auf den letzten Seiten. Ein Wort wie „athletisch“ findet man dagegen mit höherer Wahrscheinlichkeit zu Beginn eines Buches. Solche Begriffe, so die Erklärung, werden typischerweise benutzt, um Figuren zu charakterisieren, wenn sie dem Leser erstmals vorgestellt werden. Bei unscheinbaren Wörtern bringt die Analyse etwas weniger erwartbare Einblicke hervor: Das Wort „oder“ zum Beispiel steht auf den ersten Seiten deutlich häufiger als auf den letzten: womöglich, weil es benutzt wird, um Handlungsalternativen der Figuren zu benennen – und die nehmen im Laufe der Geschichte ab.

Illustration: Enrico Nagel

Dieser Text wurde veröffentlicht unter der Lizenz [CC-BY-NC-ND-4.0-DE](https://creativecommons.org/licenses/by-nc-nd/4.0-de/). Die Fotos dürfen nicht verwendet werden.

Daten | Datenschutz