

MEDIEN

BIG DATA

Die Erfahrungen von vielen

Andreas Lorenz-Meyer



Zu viel Information erstickt die Information? Im digitalen Zeitalter stimmt das nicht mehr. Neue Techniken ermöglichen es, die online anfallenden großen und unstrukturierten Datenmengen auszuwerten.

Das weltweite Datenvolumen nimmt rasend schnell zu. Zwischen zwei und drei Zettabyte soll es 2012 betragen haben. Zettabyte - ausgeschrieben ist das eine Eins, gefolgt von 21 Nullen. Fragt sich, was mit dem immer weiter wachsenden Datengebirge anzufangen ist. Die Antwort könnte Big Data heißen: das Analysieren großer, unstrukturierter Datenmengen in hohem Tempo.

Doch welchen Nutzen haben solche Auswertungen? Donald Kossmann vom Institut für Informationssysteme an der Eidgenössischen Technischen Hochschule Zürich sagt, damit könnten Erfahrungen analysiert werden. Die klassische Informatik entwickle Programme „nur“, um Prozesse zu automatisieren. Big Data verbessere diese Methode, indem es die Erfahrungen anderer nutzbar macht. Etwa bei der Wartung einer Maschine. Ob diese demnächst ausfällt, kann mit Hilfe der Daten anderer Maschinen



vorhergesagt werden. So kommt es nicht zu einem unerwarteten Ausfall.

Siemens setzt auf dieses Frühwarnsystem. Die Turbinen der Kraftwerke sollen per Big Data gewartet werden. Und das Smart Data Innovation Lab in Karlsruhe nutzt „intelligente Daten“, um das Verhalten von Besucherströmen vorherzusagen. Dazu werden die Erfahrungen früherer Großveranstaltungen ausgewertet.

Das nächste Buch, das nächste Kind

Auch der Online-Handel versucht, Ordnung ins Datenchaos zu bringen. Die Kaufempfehlungen von Amazon

basieren auf einem geheim gehaltenen Algorithmus, der die Erfahrungen von vielen durchleuchtet. Dann erscheint ein beiläufiger Hinweis auf dem Bildschirm: Kunden, die dieses Buch gekauft haben, interessieren sich auch für jenes Buch oder jenen Gegenstand. Amazon verkauft die Daten auch an andere. Kinesis heißt ein neuer Dienst, der stündlich mehrere Terabytes analysieren kann. Klickströme oder Finanztransaktionen werden in Echtzeit verarbeitet, also in dem Moment, in dem sie anfallen.

Amazon bereitet schon den nächsten Schritt vor. Big Data soll zu einem „vorausschauenden Versand“ führen. Dabei werden Produkte schon vor der

Bestellung geliefert. Und zwar genau an den Ort, wo sie laut Big Data auch gekauft werden. Der Konzern traut seinen Algorithmen offenbar den Blick in die Zukunft zu.

Big Data entdeckt Zusammenhänge, auf die niemand von allein käme. Der US-Einzelhandelskonzern Walmart stieß auf ein unerklärliches Konsumphänomen. Ein sich nähernder Hurricane steigert - warum auch immer - das allgemeine Verlangen nach Pop Tarts, einem süßen Frühstücksgebäck. Mit diesem Wissen wird Umsatz gemacht: Bei unruhiger Wetterlage liegen die Pop Tarts gleich am Eingang aus.

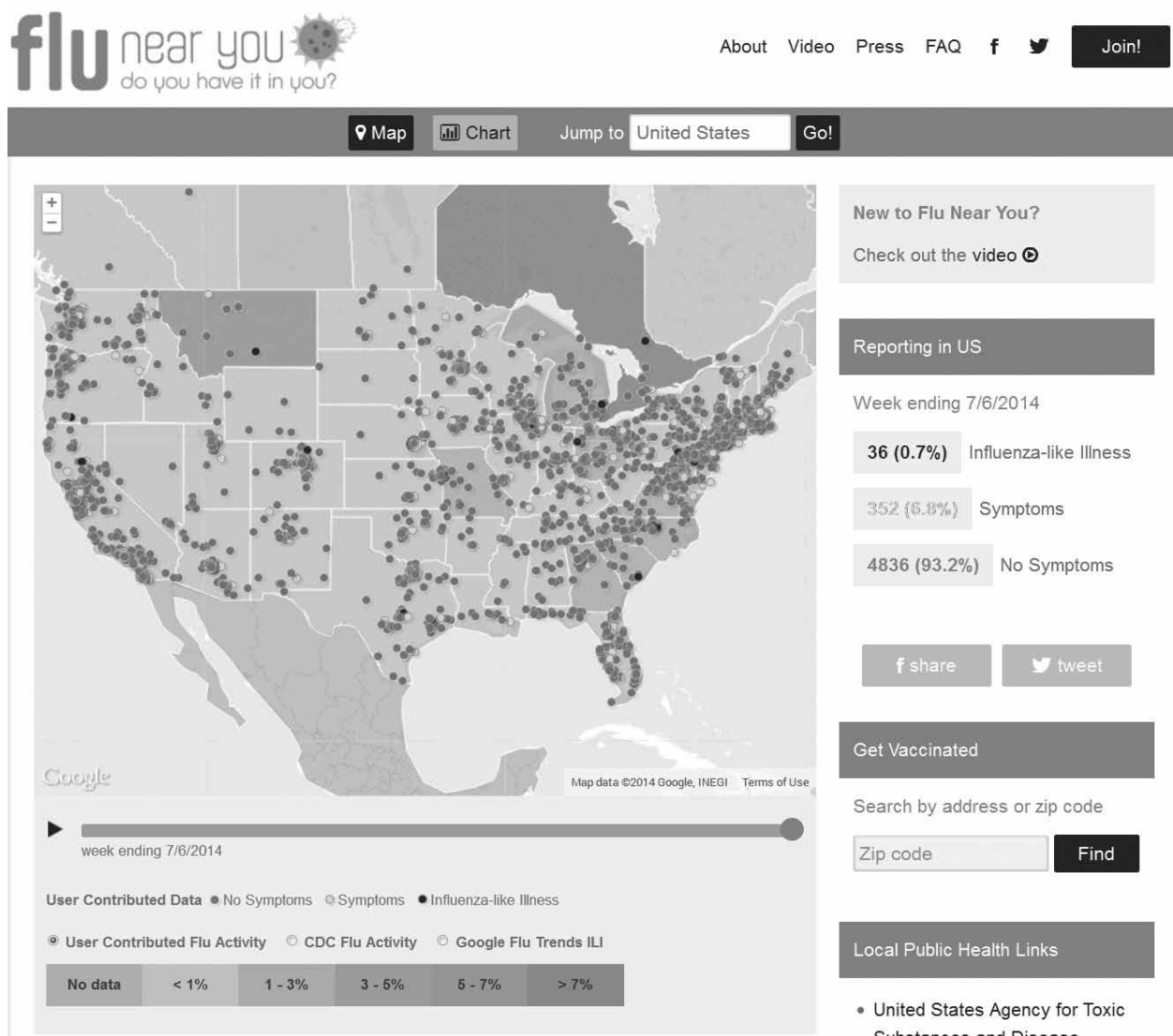
Self-defeating prophecy

Der datengestützte Verkauf kann allerdings zu weit gehen. Ein Einzelhändler erkannte anhand von Kaufmustern sehr früh die Schwangerschaft junger Kundinnen. Diese kauften größere Mengen unparfümierter Lotionen oder deckten sich mit Zink- und Magnesiumtabletten ein. Der Händler schickte daraufhin - ungefragt - die „passenden“ Geschenkgutscheine, etwa für Windeln.

Big Data führt zu einer „personalisierten Medizin“. Die Idee: Der Arzt

© 2013 SCHWARTZ ET AL. / CC BY 3.0

Nicht nur Computer, auch Menschen sollen die aus Big Data extrahierte Information verstehen können - dabei hilft eine grafische Aufbereitung. Auf Seite 12 die aus 700 Millionen Facebook-Einträgen gewonnenen Erkenntnisse über die Lieblingswörter von Frauen (oben) und Männern (unten). Der Original-Artikel „Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach“ trägt die DOI-Referenz 10.1371/journal.pone.0073791. Rechts ein Screenshot des Projekts „Flu near you“, das unter anderem auf Google zurückgreift, um über Grippewellen zu informieren.



greift auf Patientendaten aus aller Welt zu, um schnell - nach wenigen Minuten - die bestmögliche Therapie zu finden. Mit Big Data stehe ihm ein viel größerer Erfahrungsschatz zur Verfügung, erklärt Kossmann. Dies sei besonders wichtig bei seltenen Krankheiten oder außergewöhnlichen Kombinationen von Krankheiten.

In Potsdam nimmt die Big-Da-ta-Medizin schon Gestalt an. Der Hochleistungsrechner am Hasso-Plattner-Institut kann genetische Veränderungen „in Echtzeit“ analysieren, indem er die Ergebnisse von medizinischen Datenbanken aus aller Welt heranzieht. Genomdaten können so „binnen Sekunden“ analysiert werden. Die Ärzte verlieren keine Zeit mehr mit wochenlangen Recherchen.

Auch Internet-Konzerne, deren Kernkompetenz ja nicht gerade die Gesundheit ist, versuchen sich an medizinischen Prognosen. Google sagt, ausgehend von der regionalen Häufigkeit bestimmter Suchbegriffe, Grippewellen vorher. Bei der Methode sind allerdings Zweifel angebracht. Im letzten Winter schätzte „Flu Trends“ die Zahl der Kranken in den USA viel zu hoch ein: 10 Millionen sollten es sein, 5 Millionen waren es.

Kossmann erklärt, wie es zu der Fehlschätzung kam: „Wenn Google eine große Grippewelle voraussagt, dann sind die Leute vorsichtiger und waschen sich häufiger die Hände. Vorhersagen verändern also das Verhalten. Dadurch könnte eine Grippewelle eventuell sogar verhindert werden.“

Gefahren und Grenzen

Wo bleibt bei Big Data der Datenschutz? Das Unabhängige Landeszentrum für Datenschutz in Kiel befürchtet „massive Verletzungen“ informationeller Grundrechte. Alles wäre mit allem kombinier- und dann auswertbar: Angaben über Finanztransaktionen, Bonität, medizinische Behandlung, privaten Konsum, Berufstätigkeit. Werbekampagnen verkaufen das „Heilsversprechen einer besseren Informationsgesellschaft“.

Häufig heißt es auch beruhigend, die erhobenen Daten würden anonymisiert. Aber sind sie dann wirklich nicht mehr zuzuordnen? Ein belgischer Mathematiker führte ein Experiment durch. Er bekam anonymisierte Mobilfunkdaten, die nur verrietten, wann und wo sich ein Handy mit einem Funkmast verbunden hatte.

Anhand einer Formel, die der Forscher entwickelte, konnten die Bewegungsmuster aber doch zugeordnet werden. Vier zufällige Datenpunkte reichten, um die Identität einer Person mit 95-prozentiger Genauigkeit zu bestimmen.

Anonymisiertes kann also auch wieder entanonymisiert werden. Deswegen darf sich Big Data auch nie auf Einzelne beziehen, sagt Kossmann. Die Ergebnisse müssten immer aggregiert, also zu größeren Einheiten zusammengefasst werden. So bleibe der Datenschutz gewahrt. Fragt sich, wie diese Grenzziehung - allgemeine Aussagen ja, Einzelzuordnungen nein - kontrolliert wird.

Kossmann warnt vor einer „blinden Nutzung“. Datenmassen lieferten zu jeder Frage eine Antwort, selbst zu einer unsinnigen. Etwa: Ist in einem Autobahnstau die linke oder die rechte Spur schneller? Je nach Ergebnis wechseln natürlich alle auf die schnellere Fahrbahn, die bei erneuter Auswertung die langsamere wäre. Kossmann: „Erfahrung aus der Vergangenheit ist nicht immer geeignet, etwas vorherzusagen. Wir müssen stets den gesunden Menschenverstand einschalten. Auch wenn die Verwendung von Big Data so bequem erscheint.“

Big Data wird als Revolution bezeichnet, als Währung des 21. Jahrhunderts. Millionen neue Arbeitsplätze sollen entstehen. Technisch stünde dem Boom nichts im Weg. Erfahrung kann mit immer ausgefeilteren Algorithmen immer genauer automatisiert werden. Allerdings fällt es digitalisierten Gesellschaften noch schwer, einen Nutzen aus den gewaltigen Datenbeständen zu ziehen, die sie laufend produzieren. Die Kunst, sich im Datendickicht zurechtzufinden, muss erst gelernt werden.