

2. Januar 2013 08:38 "Big Data"

Wenn Daten sprechen

Von Bernd Graff

Besteht zwischen Bewegungsprofil und Schuhgröße womöglich ein Zusammenhang? Kann man durch die Auswertung von Tweets drohende Verarmung und Epidemien prognostizieren? Unter dem Stichwort "Big Data" denken IT-Firmen darüber nach, wie sie mehr Nutzen aus ihren unendlichen Datenbeständen ziehen können. Noch hat keiner eine Lösung - aber der Druck ist groß.

Was hätte Carolus Magnus, der König des Fränkischen Reiches und seit dem 25. Dezember 800 auch Römischer Kaiser mit dem Beinamen "Der Große", was also hätte dieser gewaltige Potentat des Frühmittelalters getan, wenn ihm seine Untergebenen aus dem Boden sprudelndes Erdöl gezeigt hätten?

Der Karolinger würde vielleicht davon gewusst haben, dass man schon lange vor seiner Zeit in Mesopotamien Bitumen entdeckt hatte, das man zum Abdichten von Schiffsplanken nutzte. Vielleicht wusste er auch, dass die römische Armee Erdöl als Schmierstoff für Achsen und Räder gebrauchte. Ja, er könnte auch das "griechische Feuer" gekannt haben, das im Byzantinischen Reich als Kriegswaffe eingesetzt wurde. Damit waren Flammenwerfer gemeint, die brennendes Erdöl verschossen. Aber sonst? Karl der Große kannte weder Techniken zur Destillation und Raffinierung des Erdöls, noch hatte er dafür einen Bedarf. Er hätte dem übel riechenden Zeug darum keine weitere Beachtung geschenkt.

Keine Technik und keinen Bedarf, folglich kein Interesse. So geht man heute nicht mehr mit Rohstoffen um. Es beginnt schon damit, dass man etwas, das in großer Menge vorliegt, vorsorglich zur Ressource erklärt und sogleich Bedarf anmeldet. Auch wenn die Techniken zur Gewinnung und Ausbeutung des Urstoffes vielleicht gerade erst einmal ansatzweise entwickelt sind. Das ist die Ausgangslage der gegenwärtigen Nervosität, die das Schlagwort, nein: die Losung: "Big Data" auslöst. Man weiß noch nicht, was man mit Daten sonder Zahl anfangen soll, aber es wird groß, so groß. Das ist Hoffnung wie Versprechen. So tönt es.

So tönt es von Microsoft, Apple, Google, IBM, EMC, Facebook, Amazon, Hewlett-Packard, Hitachi, Netapp und vielen anderen. "Big Data" ist nach "Mobile" und "Cloud", nach "Echtzeitanalyse" und "Personalisierung" das *next big thing* der IT-Branche. Und es ist kein Zufall, dass gerade die größten und erfolgreichsten der börsennotierten Unternehmen jetzt laut über das Schürfen, also Auswerten ihrer Datenbestände nachdenken, welche in jeder beliebigen Sekunde über Computer, Handys, Tablet-PCs und Netzwerkrechner zusammengetragen werden und stetig wachsen.

Hinzu kommen Behörden, die den öffentlichen Raum mit Videokameras überwachen, aber die Bildermengen nicht mehr bewältigen können, weil sie nicht Geschehnisse wahrnehmen, sondern immer dieselbe Echtzeit in, sagen wir, 48 verschiedenen Einstellungen. Diese Behörden und Unternehmen verfügen also jetzt schon über unfassbare Mengen an Nutzer- und Verkehrsdaten, aus Fernspreck-Verbindungen, Web-Traffic und Sozialen Netzen. Dazu kommt, was jede Logistik-

Software über Warenbestände und Zulieferungen wie Kundenbestellungen aufnimmt.

Doch es ist eins, wenn ein einzelnes Unternehmen jährlich "Big Data" im Umfang von Petabytes, eine Zahl mit 15 Nullen, erfasst und weltweit Zettabytes, eine Zahl mit 21 Nullen, angehäuft werden. Ein anderes ist es, aus diesen Daten etwas zu machen. Sie zu lesen, zu verstehen, sie zu analysieren. Und das gelingt nicht. Noch nicht.

Der Buzz um Big Data ist also erst einmal schiere Verzweiflung an der großen Zahl, verkauft als Freude. Es ist das ungläubige Staunen auch darüber, dass sich die Serverfarmen mit ihren riesigen Daten-Reservoirs wie von selbst füllen. Angeblich verdoppelt sich das Gesamt-Datenvolumen alle zwei Jahre. Das Heu ist also da, aber man will die Stecknadeln darin.

Allein: was sind das für Nadeln? Verdächtige? Konsumenten? Kranke? Und wie will man sie finden? Denn klassische Datenbank-Anwendungen durchdringen den unstrukturierten Wust nicht mehr, auch Rechner, die schnell genug wären, damit zu operieren, sind noch rar. Insofern spricht Sabine Bendiek, eine Analystin von EMC, einer US-Firma, die sich auf das Datendurchforsten spezialisiert hat, auch von "toten Daten". 80 Prozent der Digitalbestände seien für herkömmliche Datenbankabfragen unbrauchbar. "Big-Data"-Anwendungen sollen sie erst "zum Leben erwecken".

Gemeint ist damit, dass man ja nicht nur das Web-Geschwätz, die Suchanfragen, Bilder und Videos erfasst hat, sondern auch Daten aus Mikrosensoren, die in Industrie-Apparaturen und Autos, in Satelliten und Kommunikationsgeräten verbaut sind. Diese messen und registrieren Standorte, Bewegungen, Vibrationen, Temperaturen, Feuchtigkeit und sogar die chemischen Veränderungen in der Luft. Wenn man das nun alles zusammenbringen und auswerten könnte, wenn man Verhalten, Umwelt und Kommunikation auf irgendeinen gemeinsamen Nenner bringen könnte, wenn man, um im Bild zu bleiben, die Raffinierung des Rohstoffes im Griff hätte - dann wären in Zukunft die Verarbeitung natürlich gesprochener Sprache, Muster-Erkennung und Profilierung von Daten, ja die Selbstoptimierung von Maschinen durch eigenständiges Lernen denkbar. Sagt man. Glaubt man.

Gary King, der Direktor des Harvard-Instituts für Quantitative Soziologie, schwärmt jedenfalls schon: "Wir sind noch am Anfang, aber schon gut unterwegs. Doch dieser Marsch der Quantifikation, den uns die Big Data ermöglichen, wird durch Regierungen, Wissenschaft und Business hindurchfegen. Nichts wird davon unberührt bleiben." Soll man sich also freuen?

Chris Anderson, der Chefredakteur des *Wired*-Magazine, hat in Anbetracht der Daten schon vor vier Jahren "das Ende der Theorie" ausgerufen: "Dies ist die Welt, in der Big Data und angewandte Mathematik jedes andere Erkenntnis-Werkzeug ersetzen. Weg mit jeder Theorie zum menschlichen Verhalten - von der Linguistik bis zur Soziologie! Vergesst Taxonomie, Ontologie und Psychologie! Wer weiß schon, warum Menschen sich so und nicht anders verhalten? Der Punkt ist, sie tun es, und wir können es mit beispielloser Genauigkeit messen und erfassen. Wenn wir nur genug Daten haben, sprechen sie für sich selber."

Bevor man nun anfängt, solche Visionen von der Datenauswertung in Echtzeit zu wägen, muss man sich klar machen, wer gerade von Big Data schwadroniert. Es sind eben Unternehmen, Analysten, die diesen Unternehmen zuarbeiten, wissenschaftliche Einrichtungen und Regierungsstellen. Sie alle stehen unter hohem

Rechtfertigungsdruck, sie müssen Aktionäre bei der Stange halten und Kunden akquirieren, sie müssen Drittmittel einfordern und ihr Behördentum sinnvoll und effektiv erscheinen lassen. Dafür ist nichts besser geeignet als der Gesang von Fortschritt und neuen Marktchancen. Wenn man nun den Wust aus 80 Prozent bislang unbrauchbarer, unstrukturierter Daten zu Gold erklären kann, umso besser. Ein bisschen hippe Zukunft muss man ja überall im Köcher haben, sonst gilt man als rückständig.

Big Data ist also die andere Seite jener Medaille, auf der Artificial Intelligence steht. Denn es ist klar: Wenn vorhandene Datenbanklösungen nicht mehr reichen, um Muster im Kundenverhalten, um Verdächtige auf Bahnhöfen zu erkennen und gründlichere Profile zu erstellen, dann setzt man nun auf den Algorithmus an sich. Big-Data-Lösungen wollen darum nicht mehr finden, wonach Menschen gesucht haben. Sie wollen mit automatisierter Analyse jene Muster (Personen, Verhaltensformen, Interessen, Gesinnungen) aufzeigen, an die bislang niemand gedacht hat. Wenn IBM seinen Kunden verspricht, jetzt würden "Fragen beantwortet, die bislang unerreichbar waren", dann untertreibt die Firma noch.

Man will Antworten geben auf Fragen, die bisher noch niemand gestellt hat, weil niemand auf die Idee gekommen wäre. Welcher Zusammenhang etwa bestehen könnte zwischen dem Musikgeschmack eines männlichen Weißen, seinem Bewegungsprofil in den Städten, seiner Schuhgröße und dem Wetter - ein fiktives Beispiel. Google, Facebook, Amazon gelten als Pioniere, die mit ihren "intelligenten", von keiner Menschenhand mehr zugeschalteten Formen der personalisierten Werbung schon leben, was andere gerade erst zu träumen beginnen: Die Firmen programmieren den Algorithmus, danach halten sie nur noch ihre Geldsäcke zum Scheffeln der Erlöse auf.

"Global Pulse", eine Initiative der Vereinten Nationen, versucht, über die Auswertung von Suchanfragen bei Google, etwa nach "Grippe" und "Grippesymptome", die Entstehung von Epidemien mitstenografieren zu können. Über die Verkaufszahlen von Prepaid-Karten für Handys und die sogenannte "Sentiment-Analyse" in Postings bei Twitter und Facebook will man den Verlust von Arbeitsplätzen, drohende Verarmung und das Auftreten von Seuchen in ganzen Regionen prognostizieren lernen, um rechtzeitig eingreifen zu können. Diese Postings seien "Digitale Rauchsignale". Denn, so eine Selbstaussage der Abteilung: "Big Data hält nicht nur Antworten darauf bereit, welche Produkte Menschen gerne kaufen und welche Dienstleistungen sie in Anspruch nehmen möchten, sondern auch darauf, wie sie mit Stress, Arbeitslosigkeit und Katastrophen umgehen."

"Vielleicht ist es nicht beabsichtigt, aber Big Data hat einen sinistren Beigeschmack", sagt Fred R. Shapiro, der Herausgeber des "Yale Book of Quotations". Das kann man wohl sagen. Denn die neue Hyperstatistik kann ja auch als digitale Rasterfahndung begriffen werden. Auch wenn es bei Big Data nicht in erster Linie um die Erfassung und Identifizierung Einzelner geht - "Namen sind Weißes Rauschen (Noise) in den Daten", sagt man laut *NY Times* bei Google -, so weiß doch niemand, in welchen Kontexten er unter dem Mikroskop des Algorithmus demnächst auftauchen wird: als Konsument, als Kranker, als Verdächtiger.

Tatsächlich hat das Datenmaterial Potenzial. Denn das Sammeln im großen Stil hat eben erst begonnen. Die Daten halten ewig. Und wer weiß, was man in zehn, zwanzig Jahren alles herausgefunden haben will. Die Datenberge werden ja weiter wachsen. Die Algorithmen werden genauer. Ein wenig gilt für Big Data also, was Robert Musils Mann ohne Eigenschaften einst über die Forschungen zur Ameisensäure sagte: "Was fängt man am jüngsten Tag, wenn die menschlichen

Werke gewogen werden, mit drei Abhandlungen über die Ameisensäure an, und wenn es ihrer dreißig wären?! Andererseits, was weiß man vom jüngsten Tag, wenn man nicht einmal weiß, was alles bis dahin aus der Ameisensäure werden kann?"

Heute stehen wir vielleicht noch vor einem unfassbaren Haufen Datenzeugs, mit dem keiner etwas anfangen kann. Aber wer weiß schon, was daraus werden kann - was also würde Kaiser Karl nun tun?

URL: <http://www.sueddeutsche.de/digital/big-data-wenn-daten-sprechen-1.1562758>

Copyright: Süddeutsche Zeitung Digitale Medien GmbH / Süddeutsche Zeitung GmbH

Quelle: SZ vom 02.01.2013/fran

Jegliche Veröffentlichung und nicht-private Nutzung exklusiv über Süddeutsche Zeitung Content. Bitte senden Sie Ihre Nutzungsanfrage an syndication@sueddeutsche.de.