

... mit CHRISTOPH MEINEL,  
Direktor des Hasso-Plattner-  
Instituts für Softwaresystemtechnik  
(HPI), Potsdam.



## »Keiner will der Richter sein«

**E-Mails, Fotos, Videos, Facebook-Postings: Im Internet fallen täglich sehr viele Daten an. Wie schnell können diese analysiert werden?**

Früher hätte eine Analyse von großen Datenmengen eine Ewigkeit gedauert, heute bekommen Sie in Realzeit ein Ergebnis. Nehmen Sie zum Beispiel eine Genom-Analyse: Wenn man Gendaten analysiert und untersucht, ob bestimmte Gendefekte mit bestimmten Krankheiten verbunden sind, dann sind das Vorgänge, die bei der Datenbankabfrage im Web keine drei Monate mehr dauern, sondern 20 Sekunden.

### **Was hat sich geändert?**

Die Hardware ist besser und auch billiger geworden. Zum Beispiel waren Hauptspeicher früher teuer, sie standen deshalb nur in sehr geringem Umfang zur Verfügung. Aber wenn Sie Daten analysieren wollen, müssen diese Daten vom externen Speicher, zum Beispiel einer Festplatte, in den Hauptspeicher des Prozessors transportiert und die Ergebnisse anschließend abtransportiert werden, um Platz für die Bearbeitung neuer Daten zu schaffen. Dieser Vorgang dauert hunderttausend Mal länger als das eigentliche Rechnen. Neueste Computer haben Hauptspeicher von zwei Terrabyte Kapazität oder mehr. Sie können also mit einem Mal viel größere Datenmengen vom Prozessor in Millisekunden bearbeiten lassen.

**Angenommen, ich habe einen riesigen Berg an E-Mails und Nachrichten. Wie analysiere ich diese Daten?**

Dafür gibt es unterschiedliche Verfahren. Zunächst müssen Suchkriterien für eine Vorselektion festgelegt werden, damit die Datenmenge reduziert wird. Dann könnten die verbliebenen Dokumente zum Beispiel nach bestimmten Wörtern analysiert werden, der Computer sucht dabei nach Buchstabenfolgen. Sie können auch auswerten lassen, wie wichtig ein Begriff in einem Dokument ist, indem Sie alle Wörter zählen lassen und sehen, ob Ihr Begriff häufiger im Text oder ob er in der Überschrift vorkommt. Sie können auch Personennamen mit einem bestimmten Ort verbinden, dann werden alle Personen, die mit diesem Ort in irgendeiner Weise verbunden sind oder sich dort bewegt haben, gefunden.

### **Wie gut oder fehlerhaft sind die Ergebnisse?**

Um Buchstabenfolgen zu suchen, dafür ist keine Intelligenz nötig. Allerdings wird auch nur diese Buchstabenfolge gefunden, ein Synonym dagegen nicht, daran arbeitet die Forschung noch. Der Computer würde also „Haus“ finden, „Gebäude“ oder „Hütte“ oder „Villa“ nicht. Und wenn Sie Personen mit Orten verbinden lassen, dann werden auch alle Personen erfasst, die vielleicht nur zufällig mit dem Zug durch diesen Ort gefahren sind und die im Handy den Lokalisierungsdienst nicht abgeschaltet hatten.

### **Also enthalten die Ergebnisse immer auch Fehler?**

Wie gut der Computer einen Datensatz analysiert, hängt letztlich von demjenigen ab, der das Analyseverfahren für den Datensatz entwickelt. Das ist wie in der Vergangenheit: Alle können in die Bibliothek gehen, aber die Leute kommen mit unterschiedlichen Informationen wieder raus. Die IT-Systeme können die Datenmenge nur reduzieren und in irgendeiner Weise relevante Verbindungen herausfinden. Am Ende sind es Menschen, die die Ergebnisse bewerten. Außerdem bleibt eine gewisse Zufälligkeit. Denken Sie an das Attentat in Boston, das über Video-Aufzeichnungen aufgeklärt worden ist. Wenn ich zufällig im selben Augenblick wie der Bombenleger durchs Bild gelaufen wäre, würde vielleicht auch nach mir gesucht.

**Andererseits konnte ein solches Attentat nicht verhindert werden.**

Wenn jemandem bewusst ist, dass die elektronische Kommunikation beobachtet wird, wird er Mittel und Wege suchen, sich anonym im Internet zu bewegen und E-Mails zu verschlüsseln. Gerade ein Krimineller. Auch mit der Datenanalyse wird es nicht gelingen, die Gefahr hundertprozentig zu reduzieren. Das zeigen die trotz solcher Untersuchungen erfolgreichen Angriffe. Wir wünschen uns natürlich, dass die Technik die tatsächlichen Verdächtigen findet und möglichst wenige Unverdächtige als Verdächtige qualifiziert. Dieses Interesse ist aber auf allen Seiten, auch auf Seiten der Dienste, weil es ihre Arbeit effizienter macht und Kraft spart.

**Ist dann eine solche Datenanalyse überhaupt verhältnismäßig? Viele Menschen werden überwacht, aber Attentate trotzdem nicht verhindert.**

Immerhin konnte in Boston ein noch größeres Attentat verhindert werden. Keiner will der Richter sein und sagen, lass uns das, was möglich ist, nicht machen, es wird schon gut gehen. Hinzu kommt: Die Menschen geben Google, Facebook und anderen ihre Daten, ohne sich darüber Gedanken zu machen, dass die Firmen mit diesen Daten Geld machen. Dass Sicherheitsdienste diese Daten zu benutzen suchen, liegt auf der Hand. Aber ich habe vor den staatlichen und parlamentarisch kontrollierten Einrichtungen in einer Demokratie weniger Angst als vor dem Missbrauch meiner Daten durch Kriminelle oder durch das Gewinnstreben von halbseidenen Unternehmen.

**Die Fragen stellte** MARKUS WERNING.