

# Schürfer im Datenbergwerk

Computer lernen immer besser, aus unseren Datenspuren herauszulesen, wer wir sind und was wir wollen. Eine Fähigkeit, die immer wertvoller wird: Für Unternehmen, die uns günstigstenfalls maßgeschneiderte Produkte, schlechtestenfalls maßgeschneiderte Werbung präsentieren wollen, aber auch für Wahlkämpfer und Geheimdienste. Ihnen allen gefällt die Idee des komplett berechenbaren Menschen. Es gibt aber auch nützliche Anwendungen. *Von Thomas Huber*

Wer lebt, hinterlässt Spuren. Nahezu jede Handlung eines Menschen, der sich durchs Internet bewegt, lässt sich aufzeichnen, aber auch, wer nur telefoniert, Geld abhebt oder mit der Kreditkarte zahlt, taucht automatisch in Protokolldateien und anderen Datenbeständen auf – nicht immer namentlich, aber immer mit einer nachvollziehbaren Datenspur.

Der amerikanische Journalist und Autor Stephen Baker beschreibt in seinem Buch „Numerati – Datenhaie und ihre geheimen Machenschaften“, wie wichtig diese Datenspuren für Unternehmen und Geheimdienste sind, aber auch für die Wissenschaft geworden sind. Der Titel suggeriert, diese Leute wollten mindestens die Weltherrschaft an sich reißen; im Inneren des Buchs wird das Thema dann nicht so heiß gegessen, wie es auf dem Titel gekocht wird. Die Numerati wollen, so Bakers These, nur in exotischeren Fällen unsere Seele. Meist wollen sie das Geld, manchmal die Wählerstimme – und manchmal auch die Welt verbessern.

Eine seriöse Bezeichnung für Bakers Numerati gibt es bislang nicht. Die Gewinnung von Informationen aus großen Datenbeständen wird landläufig meist „Data Mining“ genannt – also etwa „Schürfen im Datenberg“. Offiziell trägt sie den klaren Namen „Knowledge Discovery in Databases“, also ungefähr „Wissensgewinnung aus Datenbeständen“. Informatiker, Mathematiker, Psychologen und Wirtschaftswissenschaftler beschäftigen sich mit dem Thema. Ihre Forschungsergebnisse sind begehrt, denn meist geht darum, das Verhalten von Menschen vorhersagen zu können.

## WERBER Klicken Sie bitte hier!

Zu Bakers Kronzeugen gehört das Unternehmen Tacoda, das die Datenspuren von Internet-Surfern analysiert und so Informationen darüber gewinnt, wo und wie Werbung am gewinnbringendsten platziert werden kann. 20 Milliarden Einzelinformationen sammelt Tacoda pro Tag ein. Das geschieht anonym, denn bei der Art und Weise, wie die Firma das tut, wäre es erstens sehr kompliziert und zweitens sehr verboten, den Namen des Benutzers herauszufinden, der die Daten erzeugt hat. In diesem Fall ist der Name auch keine sonderlich interessante Information, denn Unternehmen wie Tacoda, von denen es viele gibt, interessieren sich für das große Ganze.

Oft genug kommt es vor, dass größere Gruppen von Benutzern genau das gleiche tun. Baker beschreibt als Beispiel einen eigentümlichen Zusammenhang: Eine bemerkenswerte Anzahl von Menschen, die auf die Anzeigen von Autovermietungen klicken, hatten zuvor Todesanzeigen in den Online-Ausgaben von Zeitungen gelesen. Warum? Kennt man amerikanische Verhältnisse, ist das leicht zu erklären: Das Land ist bekanntlich riesengroß. Wenn Freunde und Bekannte sterben, müssen die Trauern-

den nicht selten fliegen, um rechtzeitig zur Beerdigung zu kommen. Am Zielort mieten sie sich ein Auto und fahren zum Friedhof. Tacodas Analyse hat gezeigt, dass es deshalb für Autovermieter ein makabres, aber lohnendes Geschäft sein kann, für amerikanische Nutzer auf Seiten mit Todesanzeigen zu inserieren.

## FAHNDER Die Alleswissenwoller

Nicht nur eifrige Internet-Benutzer hinterlassen Datenspuren. Genau genommen genügt ein Bankkonto, ein Telefon und ein gelegentlicher Gang zum Arzt. Informationen über das Finanzgebahren von 65 Millionen Wirtschaftssubjekten in Deutschland landen bei der Schufa, der Auskunftei der kreditgebenden Wirtschaft. Und die Schufa-Akte wiederum entscheidet über Kreditvergaben, Handy- oder Internet-Verträge und anderes. Vor der Datensammelerei gibt es kaum ein Entkommen. Weil seit dem 1. Januar 2008 die Verkehrsdaten der Telekommunikation für mindestens sechs Monate gespeichert werden müssen, häuft jeder Bürger, der ein Telefon oder ein Handy besitzt, einen weiteren kleinen Datenberg auf, in dem seine Telefonverbindungen, deren Ziel, Zeitpunkt und Dauer enthalten sind.

Diese Datensammlung soll bei der Verfolgung und Bekämpfung schwerer Straftaten und von Terrorismus helfen. Besonders exzessiv wird seit den Terroranschlägen vom 11. September 2001 gesammelt. Vorreiter sind die USA, wo im Namen der Terrorbekämpfung Datenbanken von FBI und CIA zusammengeführt wurden, angereichert durch die Datensammlungen der National Security Agency (NSA). Letztere ist jener Teil der amerikanischen Geheimdienst-Infrastruktur, der sich vornehmlich mit der Beschaffung und Auswertung von Datensammlungen – sprich: mit dem Lauschen und Verstehen – befasst.

Erklärtes Ziel der NSA-Experten ist es heute, Terroristen schon anhand ihrer Verhaltensmuster ausfindig zu machen, bevor sie überhaupt zuschlagen können. Allerdings sind die Erfolgsaussichten durchaus begrenzt. Stephen Baker berichtet in seinem Buch nicht nur von den mannigfaltigen praktischen Schwierigkeiten, mit denen die Mathematiker bei den Geheimdiensten kämpfen, sondern auch von den vielen falschen Alarmen, die unbescholtene Bürger in ein schlechtes Licht rücken. Denn dass jemand sich auf der Webseite islamischer Terroristen umschaut, lässt sich feststellen. Warum er das tut, weiß man deshalb noch lange nicht.

Die Absichten und Ansichten der Menschen erfährt der interessierte Leser schon eher in Blogs, Foren, bei Twitter und bei anderen sogenannten „Sozialen Medien“, die von der Aktivität ihrer Nutzer leben. Die CIA kaufte sich in jüngster Zeit bei zwei Unternehmen ein, die sich die Auswertung genau dieser Äußerungen vorgenommen haben. Visible Technology und Attensity heißen die bei-

den neuen Beteiligungen der CIA-Firma In-Q-Tel, die Informationen aus Sozialen Medien auswerten. Diese Unternehmen arbeiten mit automatisierten Programmen, sogenannten Robotern, die Informationen aus dem Netz zusammentragen.

Bezeichnend ist, dass zu Visible Kunden auch der Software-Hersteller Microsoft gehört. Überhaupt haben Unternehmen und Geheimdienste nicht selten die selben Absichten – in diesem Fall, zu erfahren, was die Leute denken.

## FORSCHER Was das Netz denkt

Diese Dienstleistung bietet auch Nicole Merkt an. Die freiberufliche Beraterin, die sich auch „Emotionsforscherin“ nennt, hat das Netzwerk „BrandPolicy“ gegründet. Auf Basis eines Modells, das sie entwickelt hat, hat das Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme (IAIS) ein Programm geschrieben, das die emotionale Bewertung eines Unternehmens oder eines Produkts im Netz vollautomatisiert auswerten soll. „Wenn man Emotionen in Texten analysieren will, muss man bestimmen, für welche Emotion welches Wort steht“, erklärt sie. Für einen Menschen ist das eine absolut triviale Aufgabe, für einen Computer eine herkulische, denn eine Eins-zu-eins-Zuordnung gelingt selten – je nach Kontext kann ein Wort sehr verschiedene Bedeutungen haben. Es kann zum Beispiel auch sein eigenes Gegenteil bedeuten: „Mit Ironie kann das System zurecht kommen, wenn diese erkennbar ist, das heißt, wenn jemand im nächsten Satz nicht relativiert“, verdeutlicht Merkt die Grenzen der von ihr geschaffenen künstlichen Intelligenz.

Bis zu 3000 Aussagen müssten manuell, also von einem Menschen, klassifiziert werden, sagt Merkt, bis die Maschine genug gelernt habe, um allein weitermachen zu können. Merks Kollegin Melanie Gnasa, die am Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme arbeitet und an der Entwicklung der Software beteiligt ist, vergleicht diese Methode mit einem Filter gegen Spam-E-Mails – auch der muss zuerst lernen, was er aussortieren soll und was nicht.

Noch ist die Methode experimentell. Wenn sie zur Reife gelangt, könnte sich Gnasa noch ganz andere Anwendungsmöglichkeiten vorstellen: Ein solches Programm könnte zum Beispiel erkennen, wie wütend der Einsender einer Beschwerde-E-Mail beim Verfassen seiner Nachricht war, und die Priorität entsprechend festsetzen. Auch Prognosen zu politischen Entwicklungen sind möglich: „Wir hatten mal testweise vor der Landtagswahl in Hessen ein Forum analysiert, und das vorhergesagte Ergebnis kam dem Wahlergebnis erstaunlich nah“, berichtet Gnasa.

## BÜRGER Von guten Mächten

Geht es also immer nur ums Ausspähen, Durchschauen, um den grenzenlosen Wissensdurst von Konzernen, Geheimdiensten und Forschern? Nicht nur. Stephen Baker beschreibt in seinem Buch die Anwendung von Data-Mining-Methoden bei der Überwachung von Patienten, speziell in der Geriatrie, wo eine frühzeitige Diagnose sehr hilfreich sein kann, Probleme zu erkennen und zu mildern. Eins von Bakers Beispielen ist der „Zauberteppich“, ein großflächiger Gewichtssensor, der nicht nur Alarm schlagen kann, wenn der Patient plötzlich rapide abnimmt, sondern auch darauf aufmerksam macht, wenn sich ein alter Mensch plötzlich gar nicht mehr bewegt. Data Mining wird in diesem Fall beispielsweise gebraucht, um Fehler herauszuzurechnen: Das System soll keinen Alarm schlagen, wenn der Patient auf einen Schlag ein paar Kilo zu- und wieder abnimmt, weil er seine Katze ein bisschen auf den Arm genommen hat.

Und nicht zuletzt ist da der menschliche Faktor, der Baker veranlasst, die Möglichkeiten des Data Mining positiv einzuschätzen: Auch Partnervermittlungs-Agenturen versuchen, aus Daten Erkenntnisse darüber zu gewinnen, wer am besten zu wem passt. Sein eigenes Experiment mit seiner Frau geht dabei so lala aus. Sie würden leidlich zusammenpassen, sagt der Computer, der anscheinend doch nicht alles wissen kann: „Soweit ich es beurteilen kann“, schreibt Baker, „ist unsere Ehe ein beinahe mysteriöser Triumph über alle Wahrscheinlichkeiten.“