

Teraflops und Petabytes

von Tom Schimmeck

Mein erster Laptop hatte noch gar keine Festplatte, nur zwei kleine Diskettenlaufwerke. Auf dem einen lief das Betriebssystem, auf dem zweiten konnte ich Text speichern. Das ging ganz gut. Die erste Festplatte kam später, sie hatte 20 Megabyte. Das galt damals als gigantisch: 20 Millionen Zeichen.

Am Anfang war das Wort. Und das Wort ist sparsam. Die Bibel passt leicht auf eine alte Diskette. Der verwackelte Digital-Schnappschuss von Tante Käthes 80. Geburtstag schon nicht mehr. Multimedia ist gefräßig. Ein Liedchen in CD-Qualität schluckt schnell 50 Megabytes, ein Film etliche Gigabytes. Auf etwa zehn Terabytes (10 000 000 000 Bytes) wird der gedruckte Bestand der U.S. Library of Congress geschätzt. Alle amerikanischen Forschungsbibliotheken machen rund zwei Petabytes aus (eine Zwei mit 15 Nullen). Die digitalisierte Summe aller von Menschen jemals gesprochenen Worte schätzt man auf fünf Exabytes (fünf Trillionen Bytes oder Buchstaben).

Kein Wunder, dass heute jeder Aldi-Computer tausendmal mehr Speicherkapazität bietet als meine erste Festplatte. Das Datenmeer ist längst ein Ozean, nimmt allmählich intergalaktische Ausmaße an. In jeder Sekunde strömen Terabytes durch die weltweiten Datenstränge. Alles, was wir tun, produziert neue Daten. Ständig loggen wir uns irgendwo persönlich ein, um etwas zu lesen, zu überprüfen, mitzuteilen oder zu shoppen. Die wachsende Zahl von Kärtchen in der Brieftasche hinterlässt viele Spuren. Der Kauf eines simplen T-Shirts per Kreditkarte etwa bringt diverse Datenbanken in Bewegung: Die Lager- und die Kundendatei des Herstellers, unsere Konten beim Kreditkartenunternehmen und der Bank. Dazu vielleicht Werber oder Datenhändler, die sich sehr dafür interessieren, wie groß, wie alt, wie reich wir sind und welche Farben und Marken wir bevorzugen. Womöglich wurden wir beim Kaufakt auch von einer Überwachungskamera gefilmt. Und so weiter. Daten, Daten, Daten. Wer je in einer „Serverfarm“ stand, einer jener riesigen klimatisierten Hallen, wo Abertausende Rechner und Festplatten schnurren, hat ein Gefühl von den Dimensionen und vom Tempo des Informationszeitalters.

Vor vielen Jahren bereits begannen Unternehmen damit, ihre rasant anschwellenden digitalen Bestände mit neuen Methoden zu durchwühlen. Solches „Data-Mining“ dient den verschiedensten Zwecken. Große Firmen etwa vermeiden unnötige Ausgaben, wenn etwa der Ingenieur A weiß, ob und welche anderen Mitarbeiter an ähnlichen Problemen tüfteln. Eine bekannte Unterwäschefirma erforschte so, in welcher Stadt welche BH-Größen verkauft werden. Heute gibt es Software, die den bei einer Firma anrufenden Kunden schon vor dem ersten Klingeln automatisch nach Kauffreudigkeit klassifiziert. Wen das Programm als tüchtigen Konsumenten identifiziert, der rückt in der Warteschlange automatisch vor. Wer weniger emsig kauft, darf lange einer netten Melodie lauschen, bevor er aus der Leitung fliegt.

Data-Mining ist überall. Börsen nutzen solche Verfahren, um Insider-Trading auf die Spur zu kommen. Casinos erwischen auf diese Weise betrügerische Croupiers. Auch Geheimdienste sind naturgemäß ganz wild auf die neuesten Fischereitechniken für den Datenozean. Die CIA



TOM SCHIMMECK, geb. 1959, schreibt als Journalist über politische und Wissenschaftsthemen für *DIE ZEIT*, die *Süddeutsche Zeitung*, das österreichische Magazin *Profil* und die schweizerische Zeitschrift *Facts*.

investiert über einen Fonds sogar Geld in vielversprechende Data-Mining-Firmen, um möglichst nah an noch nützlicheren Produkten zu sein.

Nach dem 11. September 2001 ging der US-Regierung allmählich auf, wie bürokratisch und erratisch ihre gigantische Armada von Fahndern und Schlapphüten agiert. Wichtige Hinweise wurden verschlafen und verschlampt. Neue Strukturen und Technologien mussten her. Der „Homeland Security“-Markt wurde ein Riesengeschäft. Auch weil Firmenchefs wie Thomas Siebel von Siebel Systems den Politikern erklärten, dass die Katastrophe wohl abwendbar gewesen wäre, hätte man nur rechtzeitig ihre Produkte geordert.

Vorher gab es kaum Kontrolle. Jetzt sollte sie allerorten sein. 22 Behörden vereinte man im neuen Department of Homeland Security, 26 Datenbanken, die irgendwie mit Terrorismus zu tun hatten, sollten verknüpft werden. Das TIA („Total Information Awareness Program“) der Militärforschungs-Agentur DARPA wurde umbenannt in „Terrorism Information Awareness Program“. Ziel ist es, einem US-Ermittler per Mausklick eine möglichst große Menge von amtlichen und kommerziellen Informationen aus aller Welt an die Hand zu geben: Passanträge, Visa, Arbeitserlaubnisse, Führerscheine, Verhaftungen, Autoanmietungen und Flüge, auch Kontenbewegungen und Daten aus den Bereichen Bildung, Gesundheit und Wohnen. Anvisierte Größe der Sammlung: einige Petabytes.

Schon heute spucken moderne Überwachungssysteme wie Echelon, jenes von den USA, Kanada, Großbritannien, Australien und Neuseeland betriebene Lauschnetzwerk, das weltweit Telefonate, Faxe, Emails en gros abfängt, in jeder Sekunde Terabytes aus. Und der Strom stoppt nie. Bei der National Security Agency schufteten Zehntausende von Computerexperten, Linguisten, Mathematikern und anderen Experten daran, Relevantes aus diesen von Horchposten und Satelliten gesammelten Datenmassen herauszufiltern. Die Untergrundanlagen zwischen Washington D.C. und Baltimore sollen fast fünf Hektar groß sein. Ein Buch über die NSA trägt den schönen Namen: „The Puzzle Palace“.

Morgen, soviel ist sicher, wird die Durchleuchtung des Einzelnen noch zunehmen. Die Festplatten wachsen zügig weiter. Die Rechner werden immer schneller. Der aktuelle Superstar der Supercomputer, „BlueGene“ von IBM, am Lawrence Livermore Nuklearlabor in Kalifornien im Einsatz, läuft jetzt mit 280 Teraflops – 280 Trillionen Berechnungen pro Sekunde. Japan hat angekündigt, bis 2011 eine Maschine zu bauen, die fünfmal schneller ist als alle heute auf der Erde vorhandenen Systeme. Sie soll 10 Petaflops schaffen, 10 000 000 000 000 000 Berechnungen pro Sekunde.

Man wird sie wohl brauchen. Der Datenozean wird ja ständig weiter befüllt. Beispiel: Im Herbst 2003, zwei Jahre nach dem Anschlag in New York, führten US-Behörden etwa 75 000 terrorverdächtige Personen in ihren Datenbanken. Jetzt (Mitte Februar 2006) wurde bekannt, dass das National Counterterrorism Center inzwischen 325 000 Namen auf seiner Verdächtigenliste hat. „Wir haben Listen, die Babylisten bekommen“, meint Timothy Sparapani von der American Civil Liberties Union. „Sie vermehren sich schneller als Hasen.“

Auch die Gefahr, in diesen Informationsfluten zu ertrinken, ist groß. Schnell ist jeder Überblick perdu. Die größte Datenkollektion nützt nichts, wenn keine Intelligenz dahintersteht, kein Hirn. Ein Geheimdienstler, den ich vor Jahren einmal nach dem Erfolg des ersten deutschen Computergroßeinsatzes, der Rasterfahndung nach RAF-Terroristen, fragte, lachte nur und sagte. „Nichts. Alles Handarbeit.“

Japan will bis 2011 einen Computer bauen, der 10^{15} Berechnungen pro Sekunde schafft – man wird sie brauchen.