

Alttägliches im digitalen Zeitalter

Im Jahr 2002 lag der Menschheit erstmals mehr Information digital als analog vor – spätestens jetzt war das „Digitale Zeitalter“ angebrochen. Dies wurde von Martin Hilbert und Priscila López in ihrem Aufsatz *The World's Technological Capacity to Store, Communicate, and Compute Information* festgestellt, der im Jahr 2011 erschienen war. Und schon 2007 wurde geschätzt, dass bereits 94 Prozent der weltweiten technologischen Informationskapazität digital war – nach lediglich 3 Prozent im Jahr 1993. Einer IDC-Studie aus dem Jahr 2011 zufolge verdoppelt sich das weltweite erzeugte Datenvolumen mittlerweile alle zwei Jahre.

Als Schlagwort für diese elektronisch vorhandenen Massendaten hat sich der Begriff *Big Data* durchgesetzt. Doch für heutige Anwendungszwecke ist nicht nur das schlichte Vorhandensein dieser Datenmengen zentral, sondern vor allem ihre Auswertung. Im Allgemeinen sind diese Massendaten jedoch zu umfangreich, zu komplex, zu schnelllebig oder zu gering strukturiert, um sie mit manuellen oder herkömmlichen automatisierten Methoden auszuwerten. Mithilfe künstlicher Intelligenz und der entsprechenden Algorithmen allerdings können diese Daten strukturiert und in ihnen bis dahin unbekannte Zusammenhänge erkannt werden.

So können in der Forschung durch Verknüpfung großer Datenmengen und statistischer Auswertungen neue Erkenntnisse gewonnen werden. Beispielsweise sind in einer Studie mit rund 16000 Kindern die Zusammenhänge zwischen Übergewicht und Diabetes in den Jahren von 2006 bis 2012 untersucht worden (IDEFICS – *Identification and prevention of Dietary- and lifestyle-induced health* **EF**fects

In Children and infantS). Dabei konnte statistisch festgestellt werden, dass wer schon als Kind häufig Süßes isst, auch später häufiger zu zuckerhaltigen Lebensmitteln greift und sein Risiko erhöht, Übergewicht und Stoffwechselstörungen zu entwickeln.

Doch obwohl Big-Data-Analysen nach einer modernen Errungenschaft klingen, ist deren Prinzip nicht neu: Schon bei der Entdeckung der Ursache der Choleraepidemie in Hamburg 1892 durch Robert Koch wurden von ihm den heutigen Big-Data-Methoden sehr ähnliche Mittel eingesetzt. So konnte Koch insbesondere den korrelativen Zusammenhang feststellen, dass auf der Hamburger Seite einer Straße eine deutlich höhere Anzahl Cholera-Fälle feststellbar war als auf der anderen, der Stadt Altona zugehörigen Straßenseite. Da sich eine Seuche sicherlich nicht an politischen Gemeindegrenzen orientiert, aber eine offensichtliche Korrelation zwischen der Gemeinde und der Fallzahl bestand, führte dies zur weiteren Analyse der Daten und zur Entdeckung eines Unterschieds im Wasserversorgungssystem, der die Verbreitung des Bakteriums förderte. Dieses Beispiel zeigt eine der typischen Eigenschaften auch moderner Datenanalysen: Die schon vorhandenen Daten werden explorativ untersucht, Korrelationen in diesen entdeckt und gegebenenfalls – je nach Zielsetzung der Analyse – hinsichtlich der zugrunde liegenden Kausalzusammenhänge analysiert.

Aber auch Unternehmen erhoffen sich von Big-Data-Analysen neue Möglichkeiten zur Erlangung von Wettbewerbsvorteilen, zur Generierung von Einsparungspotenzialen und zur Schaffung neuer Geschäftsfelder. Staatliche Stellen erhoffen sich bessere Ergebnisse in der Bekämpfung krimineller und

terroristischer Delikte. Und gerade diese Aktivitäten geben zum Begriff *Big Data* der Assoziation zum Begriff *Big Brother* großen Raum, dem Diktator des Überwachungsstaats in George Orwells Roman *1984*, den er bereits im Jahr 1949 veröffentlichte.

Insbesondere umfassen diese Massendaten mittlerweile auch Bereiche, die als privat gelten: Der Wunsch der Industrie und bestimmter Behörden, möglichst freien Zugriff auf solche Daten zu erhalten, sie besser analysieren zu können und die gewonnenen Erkenntnisse zu nutzen, gerät dabei unweigerlich in Konflikt mit geschützten Persönlichkeitsrechten der Einzelnen. Klassische Anwender sind derzeit vor allem die Betreiber sogenannter sozialer Netzwerke und von Suchmaschinen. Die Erfassung und Analyse der hier anfallenden großen Datenmengen ist heute alltäglich.

Was wirklich mit dem Begriff *Big Data* zusammenhängt, welche Möglichkeiten, Grenzen, aber auch Gefahren sich dahinter verbergen, ist trotz seiner politischen, wirtschaftlichen, sozialen und nicht zuletzt persönlichen Dimensionen den wenigsten Betroffenen bekannt. Deshalb ist die vorliegende Ausgabe von LOG IN diesem Thema gewidmet – im Anschluss an die vorhergehende Ausgabe, bei der das Thema *Künstliche Intelligenz* im Mittelpunkt stand. Prädestiniert, solche Themen aufzugreifen, ist ohne Zweifel ein fundierter Informatikunterricht. Mit den in diesem Heft vorliegenden Beiträgen, vor allem mit den Unterrichtsbeispielen soll gezeigt werden, wie ein solcher allgemeinbildender Informatikunterricht aussehen könnte, in dem auf aktuelle Themen eingegangen wird.

Andreas Grillenberger
Bernhard Koerber