

Das grosse Schürfen



Umgangssprachlich wird nach Gold geschürft oder gedankliche Tiefe angedeutet. Die Informatik spricht von *Data Mining* und meint damit die maschinelle Bearbeitung grosser Datenmengen nach statistischen Zusammenhängen. Aus der Werkzeugkiste der Algorithmen bedienen sich viele. Suchprogramme und Rechtschreibhilfen gehören zum Alltag, ebenso wie Handschriftenerkennung und Kundenprofile oder Kryptologie für medizinische Berichte und Steuerverfahren. Die Konsumenten, also wir alle, gehen recht unbekümmert damit um, so lange der digitale Bergbau nützliche und angenehme Dinge ans Tageslicht befördert. Nur selten stören datenschützerische Bedenken die behagliche Verbraucherruhe, etwa wenn Google ungefragt unser Strassenleben filmt oder mit höchster Auflösung die Arbeit verrichtet, die sonst Drohnen vorbehalten bleibt. Die mediale Aufregung dauert nur kurz, denn der Unterhaltungswert ist gering und das nächste Angebot elektronischer Spielzeuge unwiderstehlich.

Wer tief schürft bringt vielleicht auch Neues zu Tage. In den Abraumhalden der Metadaten tummeln sich neue Berufe. Rechnen statt lesen heisst die Devise, seit *Google Book Search* fünfzehn Millionen Bücher digitalisiert hat. Dass ein analytisches Textprogramm das Pseudonym der Harry-Potter-Autorin knackte, war immerhin eine Sommerlochmeldung wert. Statistische Stilanalysen haben eine neue Anwendung «Kulturomik» ins Leben gerufen. Eine Mischung von Kultur-, Sprach- und Wirtschaftswissenschaft, die im Korpus der Wort- und Satzgebirge nach verborgenen Trends sucht. Computerlinguistiker und Literaturinformatiker verfassen Dissertationen. Sie wollen die Sprache mit den mathematischen Methoden der Genomik entschlüsseln, daher die Bezeichnung «Culturomics», die in Analogie den Anspruch beinhaltet, das Erbgut aller Kulturen offenzulegen. Wer kann schon Millionen von Büchern lesen? Das *distant reading* grosser Korpora ergänzt das *close reading*, die mühsame Lektüre und Interpretation von Einzelwerken. Algorithmen ersetzen die subjektive Auslegung durch berechenbares, quantitatives Suchen. Der Masterabschluss in Sprachtechnologie an der ETH entspricht dem zeitgemässen Jobprofil der Bibliotheken und Verlage. Er oder sie erfüllen die Anforderungen der Sozio-Psycho-, Neuro- und Patholinguistik, der Computer-Forensik und forschenden Datenfiltrierung. In allen Fällen werden riesige Textmassen eingescannt und nach semantischen, lexikalischen und syntaktischen Mustern ausgesiebt. «Wie man eine Million Bücher nicht liest» [1] ist eines der jüngsten Positionspapiere zu diesem Vorgehen überschrieben. Die Erkennungsdienstliche Behandlung auf Triggerwörter und Auto-

renprofile hat sich seit Jahren in zahlreichen Berufen etabliert. Für die Konsumenten bietet Google, wohl um den schlechten Ruf durch die systematische Verletzung von Urheberrechten und Privatsphäre zu verbessern, ein Geografiequiz als Abfallprodukt von Streetview gratis an [2]. Ein weiteres Programm [3] ermöglicht die statistische Analyse von über fünf Millionen digitalisierten Büchern, die zwischen 1800 und 2000 englisch, französisch, deutsch und spanisch erschienen sind. Die Analyse eines Datenbestandes von rund 500 Milliarden Wörtern gibt eine Vorstellung davon, was *Data Mining* bedeutet.

Die Textfragmente, sogenannte N-Gramme, das Ergebnis der Zerlegung in Buchstaben, Phoneme und Wörter, ermöglicht auch ein Programm wie *Loading Friendship* von Facebook. Farbige Kreise und Linien kartographieren die soziale Netzwelt der Kontaktadressen. Auch Twitter offeriert ähnliche Freundeswolken oder Cluster, kolorierte Grafiken, die genau anzeigen wer mit wem wie häufig Verbindungen herstellt. Es sind Staatsfeinde wie Julian Assange von WikiLeaks und Edward Snowden von der NSA, die einer breiten Öffentlichkeit enthüllt haben, wie effizient Schnüffelprogramme arbeiten. Dass die National Security Agency NSA Megazentralen mit Kühlwasserpumpen und eigenem Stromkraftwerk errichtet, ist bekannt. Auch dass Daten von Aussenstellen in der ganzen Welt, private E-Mails, Telefonate und Suchabfragen, verschlüsselte Mitteilungen und Daten von Online-Einkäufen gesammelt werden, konnte wissen, wer es wollte.

Das Erstaunlichste dabei ist die relative Gleichgültigkeit, mit der diese Nachrichten aufgenommen werden. Ohne die wahltaktisch bedingten Schaukämpfe wären auch diese Erkenntnisse klanglos im Sommerferienloch verschwunden. Es mag dafür verschiedene Erklärungen geben. Etwa die mediale Übersättigung oder ein politisches Desinteresse, gepaart mit einer Geschichtsamnesie, die vergessen lässt, was Diktaturen anrichten. Vielleicht ist es ein übertriebenes Sicherheitsbedürfnis, das die als selbstverständlich empfundene Freiheit leichtfertig aufs Spiel setzt. Vielleicht sind Mahner einfach nur lästig, weil sie uns daran erinnern, dass die gewohnten, konsumentenfreundlichen Technologien auch den Überwachungsstaat unaufhaltsam vorantreiben. Es trifft ja nur die anderen, denkt sich die oder der und träumt weiter den grossen Spiesserschlaf.

Erhard Taverna

- 1 <http://people.lis.illinois.edu/>
- 2 www.geoguessr.com
- 3 www.ngrams.googlelabs.com

erhard.taverna[at]saez.ch