

# Von Akrostichonalgorithmen und zementierten Geschlechternormen

*von Matthias Hagen, Janosch Schobin*

## Matthias Hagen im Gespräch mit Janosch Schobin

**Soziopolis: Herr Professor Hagen, Sie haben die erste deutsche Professur für Big Data Analytics inne. Was verstehen Sie unter Big Data?**

Matthias Hagen: Das ist ja ein Kunstwort aus der Industrie. Ich persönlich drücke mich am liebsten um eine feste Definition von Big Data. Worauf der Begriff für mich hinweist, ist, wie viel man heute theoretisch über beispielsweise Menschen - ihr Einverständnis einmal vorausgesetzt - wissen kann. Großen, digitalen Datensätzen, die heterogene Informationsbestände zusammenbringen, entnehmen wir enorm viel darüber, wie sich Menschen verhalten. An der Sache ist per se natürlich nicht viel neu. Große Datensätze gab es auch früher. Der Unterschied ist heute vor allem deren bessere Verfügbarkeit. Es ist günstiger geworden, große Datenmengen zu speichern. Maschinen merken sich menschliches Verhalten daher in einer viel größeren Detailfülle und über viel längere Zeiträume.

Aus der Dichte der quantitativen Daten folgt das eigentlich Spannende. Nehmen Sie die Bewegungsmuster von Menschen: Wie bewegen sich die Menschen durch die Stadt? Oder: Wie lesen Menschen einen Text? Darüber hatten wir auch zuvor Daten, aber die wurden in der Regel nur sehr punktuell und unter Laborbedingungen erfasst. Heutige Daten werden dagegen, wenn man so will „natürlich“, in den entsprechenden Umgebungen mitproduziert und dabei gespeichert. Aufgrund dieser drei Eigenschaften, der Detailfülle, der Dichte und Dauer kann man sich als Wissenschaftler ganz neuen Fragen stellen. Etwa: Wie verändern sich bestimmte Bewegungsmuster in einer Stadt über die Jahre? Aber das betrifft nicht nur die Daten, die Menschen hinterlassen. Maschinen in einer Fabrik können auf einmal ganz ähnlich untersucht werden. Man kann sehen: Was geschah mit der Maschine vor drei Jahren und wie hat das ihre Funktionen beeinflusst? Früher wurden die entsprechende Daten nicht in dem Maße gespeichert oder einfach nach kurzer Zeit entsorgt.

**Bei meinen Recherchen zu Ihrer Person bin ich auf einige Projekte gestoßen, über die ich mich gewundert habe. Beispielsweise haben Sie einen Algorithmus entwickelt, der einen Text in ein Akrostichon umformt,**

## **also in einen Text, bei dem die ersten Buchstaben jeder Zeile zusammen Sinn ergeben. Was hat das mit Big Data zu tun?**

Das ist eigentlich ganz naheliegend. Einer der Themenkomplexe, an denen ich arbeite, befasst sich mit der Frage, wie sich die Nutzer von Suchmaschinen besser verstehen und unterstützen lassen. Die Nutzer können wohlgerne auch Maschinen sein. Texte umschreiben gehört zu diesem Komplex dazu. Texte verfassen muss man schließlich immer wieder, und darunter fallen verschiedene Tätigkeiten. Zunächst recherchiert man, findet Sachen, stellt sie zusammen, schreibt einen ersten Entwurf des Textes, und dann möchte man den vielleicht verbessern. Dabei wiederum könnte man maschinell unterstützt werden.

Was mich umtreibt, ist jedoch nicht so sehr das vollautomatisierte Erstellen von Texten, das etwa in der Finanzindustrie oder im Sportjournalismus immer mehr Raum gewinnt. Mir geht es um eine sinnvolle Schreibassistenz. Der Akrostichonalgorithmus ist dafür so eine Art „Technologie-Demonstrator“. Dass ein Text ein Akrostichon ergeben soll, ist einfach eine Vorgabe, die die Form des Textes in gewisser Weise beschränkt. Um die Bedingung einzuhalten, müssen Sätze umgestellt und Wörter ersetzt werden. Das Ziel ähnelt im Hinblick auf seine Umsetzung sehr dem Wunsch, der Text möge fünfzig Wörter kürzer oder einfacher zu lesen sein. Big Data hängt mit der Problematik insofern zusammen, als wir für das Umstellen der Sätze nicht einfach nur grammatische Regeln in den Rechner programmieren, sondern auf die Auswertung großer Datensätze zurückgreifen, um passende gebräuchliche Formulierungen sowie Wortersetzungen zu identifizieren.

Im Kontext der „Schreibassistenz“ haben wir etwa Personen gebeten, mit einer Suchmaschine über ein bestimmtes Thema zu recherchieren und uns dann ein Essay darüber zu schreiben. Wir haben dabei herausgefunden, dass es grob gesprochen zwei Typen gibt: Die einen sammeln alles am Anfang und schreiben dann ihren Text. Die anderen sammeln die Informationen fortlaufend, während sie bereits schreiben. Wir haben auch Netspeak, eine eigene Suchmaschine, entwickelt, auf die wir ziemlich stolz sind. Sie können dort beispielsweise englische Phrasen eingeben und einen Platzhalter für ein Wort definieren: Netspeak sagt Ihnen dann auf der Basis von Webhäufigkeiten, welche Formulierungen an dieser Stelle gebräuchlich sind. Das ist ganz praktisch, wenn man sich etwa nicht sicher ist, welche Präposition vor ein Wort gehört: Sagt man eher „I was at the schoolyard“ oder „I was on the schoolyard“? Wir beobachten die Nutzung unserer Suchmaschine zudem, um die empirischen Effekte der Suchmaschinenverwendung zu erforschen. Eine These lautet, die Internetnutzung führe dazu, dass sich die Leute immer weniger merken. Wir schauen daher, ob sich bei unseren Nutzern Lerneffekte zeigen oder ob sie, grob gesagt, immer wieder das Gleiche suchen.

In Zukunft werden wir uns verstärkt mit Argumenten in Texten befassen. Wir wollen herausfinden, wie man Argumente in einem Text erkennt und diese sortiert, etwa in Pro und Kontra. Ob wir das hinbekommen, weiß ich noch nicht. Aber auch

dabei benötigen wir wieder ein Werkzeug, das beim Sortieren und Verarbeiten von Information in Textform behilflich ist. Nutzer im Bereich Text verstehen und unterstützen sehe ich als einen wichtigen Teil meiner Forschung; vermutlich steht das aktuell sogar an erster Stelle.

## **Und was machen Sie an zweiter Stelle?**

Eine zweite Forschungslinie, mit der wir uns in meiner Arbeitsgruppe befassen, hat mit dem zu tun, was in den letzten Jahren vermehrt unter dem Stichwort „Industrie 4.0“ durch die Medien gegangen ist. Wir arbeiten beispielsweise mit einer Firma zusammen, die Anlagen zur Produktion von Wellpappe betreibt und als sogenannter „Hidden Champion“ gilt. Ein wesentlicher Anteil der weltweit verwendeten Wellpappe wird in deren Anlagen hergestellt. Das klingt zunächst nach gutem, altem Maschinenbau. Aber heutzutage werden in solchen Anlagen Unmengen von Sensoren verbaut. Wie heiß ist die Maschine an bestimmten Stellen, wie schnell drehen sich die Walzen, wie hoch ist der Druck und so weiter. Die Firma und ihre Kunden interessiert das Naheliegende: Was geht als nächstes kaputt? Kann man etwas Energie einsparen? Gibt es irgendwo auf der Welt einen Kunden, der seine Maschine mit weniger Temperatur und weniger Druck betreibt? Lässt sich etwas optimieren? Wir versuchen also, die Fernwartung und Fernsteuerung solcher Maschinen durch die konsequente Nutzung der Sensordaten zu verbessern. Das hat seine Potenziale, aber ist zumindest aus meiner Sicht keine gigantische Veränderung.

Ein dritter Bereich, in dem wir tätig sind, ist die Datenvisualisierung. Es geht dabei vor allem um neuartige Visualisierungskonzepte für Datenanalysen. Darunter fallen ziemlich heterogene Projekte. Ein Kollege beschäftigt sich zum Beispiel damit, wie man Textähnlichkeiten einfach visualisieren, wie man also Textwiederverwendungen visuell kenntlich machen kann. Die Verbreitung und die Umformulierung von Nachrichten ließe sich damit etwa darstellen. Hinzu kommt die Visualisierung der Ähnlichkeiten von Simulationsresultaten im ingenieurwissenschaftlichen Bereich. Zum Beispiel werden in der Industrie bei der Prototypentwicklung immer seltener einzelne Simulationen am Idealmodell durchgeführt. Stattdessen macht man tausende Simulationen an Modellen mit zufälligen Abweichungen, um schon im Vorfeld Fehlerquellen und Optimierungsmöglichkeiten zu entdecken. Da wäre es häufig ganz praktisch, über Instrumente zu verfügen, mit denen sich die Ähnlichkeit verschiedener Simulationen grafisch leicht erkennbar darstellen lässt.

**Sie haben ein paarmal betont, dass Ihre Nutzer auch Maschinen, also Computerprogramme, sein können. In der Soziologie wird mittlerweile seit fast vierzig Jahren über die Mensch-Computer-Beziehung und ihre Auswirkung auf die Gesellschaft debattiert. Einerseits gibt es diejenigen, die sich vor allem mit den Ergänzungsverhältnissen zwischen Menschen und „Denkzeug“ (eine Formulierung Konrad Zuses) beschäftigen. Donna**

**Haraway etwa sprach schon in den 1980er-Jahren von Mensch-Maschine-Hybriden, von Cyborgs.<sup>1</sup> Computer sind für diese Fraktion ziemlich unvorhersehbare Katalysatoren des kulturellen und gesellschaftlichen Wandels. Die andere Fraktion betrachtet in der Tradition von Karl Marx Computer eher unter dem Paradigma der realen Subsumption: Computer verdinglichen menschliches Wissen und substituieren auf diese Weise menschliche Arbeit sowie menschliche Arbeiter. Diese Auffassung scheint gerade wieder Oberhand zu gewinnen. So haben Frey und Osborne kürzlich prognostiziert, dass 47 Prozent der US-Jobs durch Computertechnologien ersetzt werden könnten.<sup>2</sup> Big-Data-Technologien spielen bei diesem Szenario eine wichtige Rolle. Wie sehen Sie das?**

Mein Eindruck ist, dass wir uns sehr stark auf eine Zukunft zubewegen, in der nur noch computerassistierte Menschen, oder Mensch-Maschinen-Hybride, wirklich handlungsfähig sind. Nehmen Sie das wissenschaftliche Arbeiten. Kein Mensch schafft es, die gesamte wissenschaftliche Literatur in einem bestimmten, etwas weiter gefassten Bereich noch zu überschauen. Dafür brauchen wir solche Hybride. Was die Abschaffung der Menschen in der Industrie angeht: Bei bestimmten Tätigkeiten wird das vermutlich so kommen. Wenn selbstfahrende Fahrzeuge erst einmal akzeptiert sind, braucht man keinen mehr, der sie professionell fährt. Oder nehmen Sie Übersetzungen, das liegt näher an dem, womit ich mich beschäftige. Mittelfristig werden für bestimmte Textarten keine menschlichen Übersetzer mehr gebraucht werden. Vielleicht wird es bei Romanen weiterhin sinnvoll sein, Menschen einzusetzen, aber bei technischen Übersetzungen oder einfachen Dolmetschertätigkeiten sicherlich nicht. Ich denke, dass wir dafür bald keine Menschen mehr brauchen. Ich habe natürlich die Analysen von Frey und Osborne gelesen, aber ich bin kein Sozialwissenschaftler und kann die Qualität der Prognosen schlecht beurteilen. Ich hege aber wenig Zweifel daran, dass vergleichsweise einfache Arbeiten automatisiert werden, es sei denn, wir als Gesellschaft verhindern diese Automatisierung.

**Mensch-Computer-Hybride können dadurch beeinträchtigt werden, dass ihr Computeranteil relativ autonom agiert. Der Techniksoziologe Werner Rammert spricht in diesem Zusammenhang von verteiltem Handeln in soziotechnischen Konstellationen. Unter bestimmten Umständen ist die Koordination zwischen Computeragenten und Menschen essenziell, damit Dinge gelingen. In den Human-Computer-Interaction Studies zeigten Nass und andere bereits in den 1990er-Jahren experimentell, dass schon sehr kleine Anzeichen ausreichen, damit Menschen Computeragenten im Rahmen sozialer Normen bewerten.<sup>3</sup> War die Stimme eines Computertutors männlich, wurde Lob von ihm beispielsweise freundlicher aufgenommen. Big-Data-Verfahren versprechen ja unter anderem, computergesteuerte Maschinen „sozialkompetenter“ zu machen, ihnen also die Interpretation sozialer Normen beizubringen. Können Sie mit diesem Projekt etwas anfangen?**

Wir arbeiten tatsächlich an etwas Vergleichbarem, allerdings eher algorithmisch

und nicht an Robotern. Es geht also nicht so sehr darum, dass unsere Programme soziale Normen in Handlungszusammenhängen korrekt interpretieren. Aber was für die Nutzer von Suchmaschinen immer wichtiger wird, besonders wenn es um komplexere Anfragen geht, sind Erklärungen. Wenn ich nach dem Wetter von morgen frage und eine Antwort bekomme, kann ich vielleicht noch darauf verzichten zu wissen, wie die Suchmaschine zu diesem Ergebnis gekommen ist. Ich unterstelle einfach, dass sie verschiedene Datenbanken mit meteorologischen Prognosen befragt hat und mir eine Art Mittelwert angibt.

Bei schwierigeren Fragen wird jedoch relevant, wie das Ergebnis zustande gekommen ist. Ich erinnere mich an ein lustiges Beispiel aus einem Blog: Der Blogger hatte Google gefragt: „Wie alt wird meine Katze?“ Googles Antwort lautete: „Fünfzehn Jahre.“<sup>4</sup> Woher aber will Google wissen, wie alt seine Katze wird? Welche Quellen wurden herangezogen? Wie wurden die Quellen bewertet? Welche Gründe sprechen für diese Einschätzung? An diesem Thema wollen wir arbeiten. Erklärungen würden das Vertrauen in die Suchergebnisse häufig verstärken, aber auch transparent machen, welchem Prozess sich die Information verdankt.

**Ein anderes Problem, auf das im Rahmen der Interpretation sozialer Normen durch Computer in der Soziologie aufmerksam gemacht wird, besteht darin, dass die Digitalisierung des Sozialen unter Umständen Normen perpetuiert. Um bei dem Computertutor zu bleiben: Wenn Computertutoren in der Regel Lob durch männliche Stimmen kommunizieren, dann verstärkt das unter Umständen Geschlechternormen, von denen wir uns heute emanzipieren wollen. Kevin Slavin zufolge leben wir in einer Welt, in der algorithmische Modelle, die wir zur Beschreibung der Wirklichkeit ersonnen haben, immer häufiger zu Modellen werden, die unsere Welt *mitproduzieren*.<sup>5</sup> Wie beurteilen Sie die normativen Effekte von Big-Data-Anwendungen?**

Solche Effekte gibt es bestimmt. Ich denke wieder an die Sprache. Wenn wir so etwas anklicken wie Netspeak, dann haben dessen Vorschläge einen verstärkenden Effekt. Vielleicht betreffen derartige Wirkungen nicht direkt tiefer liegende gesellschaftliche Normen, etwa Geschlechternormen, aber durchaus solche Normierungen wie Sprachregelungen oder Rechtsschreibkonventionen: Unsere Algorithmen entnehmen den Daten im Netz übliche Schreib- und Ausdruckformen. Die Nutzer orientieren sich daran, weshalb die massive Nutzung unseres Angebots dazu führen könnte, dass es einen bestimmten Sprachgebrauch fixiert. Es kann also durchaus zu Verstärkungseffekten kommen. Ich halte viele der normierenden Effekte, die dabei auftreten können, aber für Probleme, die im Rahmen einer klugen Technikfolgenabschätzung mitbedacht werden können und sollten.

Trotzdem lassen sich natürlich bestimmte Entwicklungen vorab nur schwer absehen. Im Moment ist es definitiv so, dass mir Netspeak hilft, lesbare Texte auf Englisch zu verfassen, die gebräuchliche Formulierungen nutzen. Aber vielleicht

führt diese Praxis mittelfristig dazu, dass ich schlechtere Texte schreibe, als ich hätte schreiben können, wenn ich die Suchmaschine nicht verwendet hätte. Auf solche unerwarteten Effekte können und müssen wir dann im Nachhinein reagieren, indem wir die Algorithmen anpassen oder die entsprechenden Projekte fallenlassen.

**An dieser Stelle muss ich leichte Zweifel anmelden: Algorithmen sind ja formale symboltransformierende Mechanismen. Nehmen wir einen Algorithmus, der Zeichenketten in einem Datensatz nach einem bestimmten Verfahren sortiert. Er kann potenziell auf jedes erdenkliche Sortierproblem angewendet werden und deswegen auch in vollkommen anderen Kontexten zur Anwendung kommen als denen, für die er ursprünglich entwickelt wurde. In den Medien wurde vor einiger Zeit diskutiert, dass Google nicht genügend Bilder von Frauen anzeigt, wenn man „CEO“ googelt. Wie soll man bei all den möglichen normativen Effekten selbst einfacher Such- und Sortieralgorithmen den Überblick über die Folgen neuer Big-Data-Technologien behalten?**

Ich verstehe ihren Einwand, aber ich bin weniger pessimistisch als Sie. Viele solcher Effekte sind, wie gesagt, schwer absehbar. Aber was sollen wir tun? Neue Algorithmen verbieten oder geheim halten? Oder sogar eine Algorithuspolizei einführen? Das halte ich für einen Holzweg. Man muss die Probleme von Fall zu Fall behandeln. Was man aber kritisch anmerken kann: Auf Informatikseite spielt es für die meisten Studiengänge und Kollegen, ich schließe mich da gern ein, kaum eine Rolle, die potenziellen Folgen von neuen Algorithmen früh zu verstehen. Insgesamt müsste die Beschäftigung mit Fragen der Technikfolgenabschätzung in den Informatikstudiengängen noch verstärkt werden. Zudem sollten solche Fragen in den praktischen Entwicklungsprozess einbezogen werden. Es reicht nicht, dass wir an den Unis mit ein paar Kursen versuchen, die Sensibilität der Informatiker\_innen zu erhöhen. In größeren Unternehmen etwa ließen sich die Entwicklungsteams gemischer zusammenstellen und auch Sozialwissenschaftler\_innen bei der Technikentwicklung konsultieren. Das geschieht in manchen Firmen bereits.

Insgesamt sehe ich normative Probleme von Big-Data-Technologien aber auch in einem etwas anderen Zusammenhang. Die Diskussion darüber, wem die Daten gehören, ist sicherlich wichtig. Aber aus meiner Sicht wird ein mindestens genauso brennendes Problem die Öffentlichkeit in den nächsten Jahren vermehrt beschäftigen: Die großen datengetriebenen Firmen machen riesige Experimente mit ihren Nutzern. Die ersten diesbezüglichen Skandale gab es bereits, denken Sie an Facebook. Vor einiger Zeit hat man dort heimlich und im laufenden Betrieb untersucht, wie Menschen reagieren, wenn sie mehr negative Nachrichten bekommen, während weniger positive Sachen angezeigt werden.

Das stete und vor allem nicht-deklarierte Experimentieren mit den Nutzern wirft

ethische Probleme auf. Mal überspitzt angenommen, eine Person stellt eine Suchanfrage, deren Antwort für sie über Leben und Tod entscheidet. Da sie aber gerade an einem verdeckten Experiment teilnimmt, wird ihr die dringend benötigte Antwort erst auf der zweiten Seite angezeigt. Ich übertreibe mit meinem Beispiel, aber das ethische Problem sollte erkennbar werden. Wirklichkeitsnäher könnte man an Hypochonder denken, die Krankheiten zu ihren eher unspezifischen Symptomen suchen. Würden die Suchmaschinen so optimiert, dass hypochondrische Nutzer schneller schwere Krankheiten zu ihren Symptomen finden, wäre das durchaus gefährlich. Sie würden nun allerlei medizinische Diagnostik in Anspruch nehmen, was auch den einen oder anderen Todesfall bedeuten könnte. Allerdings arbeiten große Suchkonzerne schon daran, nicht immer nur die extremsten und problematischsten gesundheitlichen Ergebnisse auf entsprechende Anfragen zu liefern, sondern das besser zu gewichten. Nicht jeder, der Kopfschmerzanfragen stellt, hat einen Hirntumor.

Nicht-deklariertes Experimentieren mit den Nutzern halte ich jedenfalls für ein großes Problem. Würden wir vergleichbare Studien an der Uni machen wollen, bräuchten wir erst die Zustimmung einer Ethikkommission und selbstverständlich die der beteiligten Personen. Im Web fragt mich bislang keiner nach meiner Experimentzustimmung. Auch so mancher Mitarbeiter der großen Konzerne ist der Ansicht, dass solche und ähnliche ethische Probleme in Zukunft auf die Tagesordnung gehören.

*Dieser Beitrag ist Teil eines Themenschwerpunkts zu Big Data. Weitere Texte finden Sie hier.*

---

#### Fußnoten

1 Donna Haraway, Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980's, in: Socialist Review 80 (1985), S. 65-108. Vgl. auch die deutsche Ausgabe einer späteren Textversion: Ein Manifest für Cyborgs. Feminismus im Streit mit den Technowissenschaften, übers. von Fred Wolf, in: Donna Haraway, Neuerfindung der Natur. Primaten, Cyborgs und Frauen, hrsg. von Carmen Hammer / Immanuel Stieß, Frankfurt 1995, S. 33-74.

2 Carl Benedikt Frey / Michael A. Osborne, The future of employment: How susceptible are jobs to computerization, Oxford 2013.

3 Clifford Nass / Jonathan Steuer / Ellen R. Tauber, Computers are social actors, in: Beth Adelson / Susan Dumais / Judith Olson (Hrsg.), the SIGCHI conference, Boston, MA, 1994, S. 72-78.

4 Vgl. Konrad Lischka, Woher weiß Google, wann meine Katze stirbt?, 17. September 2015.

5 Kevin Slavin, How algorithms shape our world, in: TED Talks 2011.

