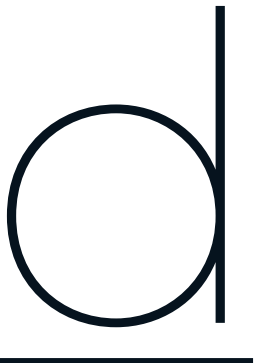


# R2-D2 AM NACHRICHTEN— TICKER

**Was der sogenannte Roboterjournalismus (nicht) kann**

Ein Gespenst geht um in der Medienbranche: das Gespenst des Roboterjournalismus. Übernehmen die Computer jetzt auch noch die Redaktionen?

TEXT Hektor Haarkötter



Der Begriff Roboterjournalismus ist irreführend und insinuiert: Man stellt sich ein humanooides Gerät wie den Roboter C-3PO aus der Star Wars-Saga vor, und häufig

werden entsprechende Artikel zum Thema auch genauso illustriert. Dabei geht es eigentlich nur um die Automatisierung der Textproduktion mithilfe von Computerprogrammen. Von Robotern also kaum eine Spur!

Nun ist Automatisierung im Journalismus nichts Neues. Gemäß DIN V 19233 ist Automatisierung »das Ausrüsten einer Einrichtung, so dass sie ganz oder teilweise ohne Mitwirkung des Menschen bestimmungsgemäß arbeitet«. Automatisierung findet auf verschiedenen Ebenen im Journalismus ebenso wie in vielen anderen Wirtschaftsbereichen schon seit langem statt, zum Beispiel im Foto- und Videojournalismus bei der Bildbearbeitung. Auch bei der journalistischen Textproduktion ist Automatisierung alltäglich. Das Kopieren von Textbausteinen, die automatische Rechtschreibkontrolle oder das Zählen von Zeichen: Das alles lassen Autoren längst von Textverarbeitungsprogrammen selbständig erledigen. Auch beim Vorgang des Schreibens und damit beim eigentlichen schöpferischen Akt lassen sich Journalisten schon länger von Computerprogrammen lenken und leiten. Wenn Überschriften von Online-Artikeln nach den Maßgaben von Klickzahlen und Google-Ranking

umgeschrieben werden, sind es automatisierte Prozesse, die die Textproduktion stimulieren. Das Schreckgespenst, das sich hinter dem Begriff Roboterjournalismus verbirgt, muss darum noch für etwas anderes stehen. Es scheint, als handle es sich um eine Art kopernikanische Kränkung des Menschen, der vom Affen abstammt, weder im Mittelpunkt des Universums steht noch beim Schach eine Chance hat, gegen Computer zu gewinnen. Jetzt übernehmen diese Computer auch noch vollautomatisch die geistige Sinnproduktion und gesellschaftliche Selbstbeobachtung in Form journalistischen Textens. »R2-D2 übernehmen Sie!« Aber stimmt das eigentlich?

Computer sind nicht zahlenverarbeitende, sondern symbolverarbeitende Maschinen und können entsprechend gut mit Buchstaben jonglieren. Experimente mit automatisch generierten Texten (NLG = Natural Language Generation) gab es bereits in der Computerfrühzeit. Das erste Programm dieser Art war vermutlich LoveLetters\_1.0, das 1952 von Christopher Strachey an der University of Manchester am Computer Ferranti Mark I programmiert worden war – auch wenn es sich dabei vermutlich eher um eine einfache

## Text|ge|ne|rie|rung

Substantiv [die]

Automatische Produktion von natürlicher Sprache durch Maschinen. Entsprechende Verfahren der Computerlinguistik werden für Chatbots oder Roboterjournalismus eingesetzt, indem aus Daten mit Hilfe von Textbausteinen Inhalte erstellt werden. Die Regeln zum Schreiben erschließen sich Computer durch selbstlernende Algorithmen.



Vertraust du Robotern?

»Phrasendreschmaschine« handelte. Eines der frühen Programme, das bereits auf eine Wissensdatenbank für die Textproduktion zurückgriff und damit sinnvollen Text produzieren sollte, war »Baseball«, das auf die Daten der amerikanischen Baseball-Liga zurückgriff. Auch literarische Texte lassen sich vom Computer herstellen: Die vermutlich erste Software zur Erzeugung stochastischer poetischer Texte wurde von Theo Lutz 1959 für den Computer ER 56 der Standard Elektrik Lorenz AG erstellt.

#### JOURNALISTISCHE TEXTE AUS DER MASCHINE

Linguistisch handelt es sich bei der automatischen Textgenese um sogenannte Permutationen. Nach bestimmten Regeln wird dabei aus einer endlichen Menge von Elementen eine bestimmte Anzahl gezogen und wechselseitig getauscht. Bei der sprachlichen Permutation ist die Menge der Elemente ein zu hinterlegendes Lexikon, und die Bauanweisungen sind die Syntaxregeln der jeweiligen Grammatik. Prinzipiell können auch mehrere Lexika verwendet und Regeln definiert werden, nach denen wechselweise aus verschiedenen Lexika Elemente zu entnehmen sind. Ein solches Lexikon kann zum Beispiel Namen von Politikern enthalten, aktuelle Ereignisse oder ein Ortsregister.

Computer-generierte Texte lassen sich mittlerweile vor allem in den Bereichen leicht erstellen, in denen die erforderlichen Fakten aus möglichst vielen Zahlen bestehen. Eine Meldung über Gewinnprognosen der Hotelkette

Marriott, die im Onlinemagazin des Wirtschaftsdiensts Forbes zu lesen ist, wirkt erst einmal wie ein üblicher wirtschaftsjournalistischer Text. Seine Besonderheit offenbart sich erst, wenn man die Unterzeile zu dem Beitrag studiert: »Earnings estimates provided by Zacks. Narrative Science, through its proprietary artificial intelligence platform, transforms data into stories and insights«. Die Daten, aus denen sich die Meldung zusammensetzt, stammen von der Wirtschaftsauskunftei Zacks Investment Research. In journalistische Form gebracht hat diese Daten aber kein Journalist, sondern das Artificial-Intelligence-Programm Quill der US-amerikanischen Softwarefirma Narrative Science. Der Artikel ist kurz, die Produktivität von Narrative Science aber enorm: Die »Text Engine« Quill kann alle dreißig Sekunden einen neuen Text generieren.

Außer der Chicagoer Textschmiede setzt auch die Firma Automated Insights mit ihrer Software Wordsmith auf automatisierten Journalismus. Das Programm wird beispielsweise von der amerikanischen Nachrichtenagentur AP eingesetzt, die pro Quartal etwa 4000 standardisierte Finanz- und Sportberichte erzeugen lässt. Bis →

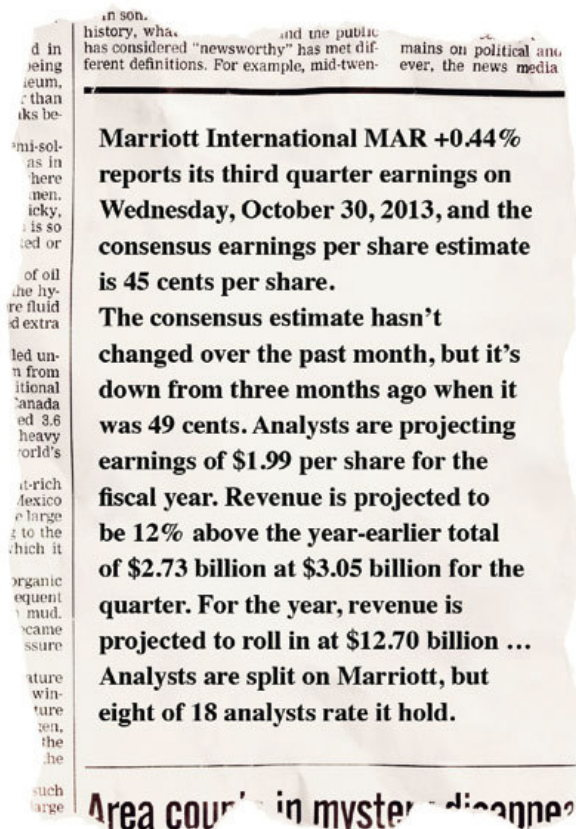
R2-D2 lässt grüßen: Immer mehr Redaktionen müssen sich an Text-Roboter gewöhnen.

#### **Per|mutat|ion**

*Substantiv [die]*

Jede mögliche kombinatorische Anordnung einer bestimmten Zahl von Elementen, in der alle Elemente verwendet werden. Sprachwissenschaftlich handelt es sich um die Vertauschung, Umstellung oder Verschiebung in der Reihenfolge von Wörtern oder Satzteilen.

2020 will AP angeblich achtzig Prozent des Nachrichtenangebotes automatisch generieren lassen. Und so sehen solche Meldungen aus:



Sieht aus wie von Menschen geschrieben: AP-Meldung einer »Text Engine«.

Auch bei Medienmarken wie Forbes, New York Times oder Los Angeles Times stammen Teile ihrer Nachrichten vom Kollegen Computer. In Deutschland hat sich die Firma Aexea mit ihrer Software Ax-Semantics der automatischen Produktion journalistischer Texte verschrieben. Die Stuttgarter Experten haben sogar eine eigene Programmiersprache entwickelt, die in Anlehnung an die Auszeichnungssprache des Internet ATML3 genannt wird und je nach Einsatzzweck von Redaktionen trainiert werden kann.

#### TURING-TEST FÜR TEXTMASCHINEN

Was aber leisten die Textmaschinen eigentlich? Die New York Times hat die Probe aufs Exempel gemacht und acht kurze Artikel online gestellt: Die Nutzer sollten herausfinden, ob die Texte vom Computer stammen oder von einem menschlichen Autor. In der Computerwissenschaft wird so etwas »Turing-Test« nach dem berühmten britischen Mathematiker und Computerpionier Alan Turing genannt. Tatsächlich lässt sich nicht entscheiden, welche Autoren-Entität die Artikel verfasst hat, Mensch oder Maschine. Über wenige Zeilen lange Nachrichten

zu datenorientierten Themen wie Sport oder Wirtschaft kommen die Textmaschinen allerdings bislang nicht hinaus. Mit strukturierten Daten können Computer eben gut arbeiten und daraus sinnvolle Sätze nach den erwähnten Permutationsregeln formen. Eine Studie an der Ludwig-Maximilians-Universität München hat ergeben, dass Leser computergenerierten Artikeln sogar eher vertrauen: Sie basieren weitgehend auf Zahlen, und das gilt als Indikator für Glaubwürdigkeit.

Etablierte Journalisten müssen zurzeit nicht um ihre Jobs bangen: Die Text-Software wird vor allem in Bereichen eingesetzt, die für professionellen Journalismus nicht relevant sind oder sich nicht lohnen. Narrative Science etwa hat die Smartphone-App Gamechanger entwickelt, mit der Trainer und Eltern die Spielverläufe der US-amerikanischen Kinder- und Jugend-Baseball-Liga erfassen und am Ende des Spiels auf Knopfdruck »a print-ready article about the game« erhalten können. Pro Spielzeit sollen so mehr als 1,5 Millionen Spielberichte erzeugt werden. Aexea kann Wettervorhersagen für 12.500 deutsche Städte (ab 2.500 Einwohner) erzeugen und damit kurze Nachrichten auf hyperlokaler Ebene verbreiten. Sogar Personalisierung von Inhalten ist möglich. Texte für eine »Audience of one«, also ein Ein-Personen-Publikum, werden die Computer-Elaborate darum auch genannt.

Warum aber scheitern Computer an der Generierung längerer Texte? Reportage und Feature beispielsweise basieren auf komplexen Strukturen und Regeln des Storytellings. An deren Implementierung forscht die Wissenschaft zwar intensiv, weil beispielsweise in der Unterhaltungsindustrie ein enormer Bedarf nach automatisch erzählten Storys bestehen soll. Doch offenbar ist der Vorgang des Geschichtenerzählens zu komplex für Maschinen, die nur Nullen und Einsen kennen. Auch die Gefahr, dass Computer künftig Kommentare schreiben und damit die politische Stimmung beeinflussen, ist gering: Computer verstehen, wie der amerikanische Sprachphilosoph John Searle bewiesen hat, gar nicht, was sie formulieren. Entsprechend können sie auch keine meinungsbasierten Texte produzieren.

Ein Journalismus der »Marke R2-D2«, also ohne Menschen, ist unmöglich. Schließlich machen auch Textprogramme Fehler, zum Beispiel fabrizieren sie Zahlendreher oder vertauschen Daten. So hat AP unlängst die Netflix-Quartalszahlen falsch verkündet: Statt des gemeldeten Kurs-Einbruchs um mehr als sieben Prozent hatte sich der Aktienwert in Wahrheit mehr als verdoppelt. Und nachdem ein Zeitungsalgorithmus der Los Angeles Times ein Erdbeben meldete, das nie stattgefunden hatte, stellte sich heraus, dass ein Mitarbeiter beim California Institute of Technology versehentlich falsche Werte in einer Datenbank korrigiert hatte. Redaktionelle Kontrolle tut also nach wie vor gut. ●