

MESSAGE

Internationale Zeitschrift für Journalismus

SOFTWARE-JOURNALISMUS

Leseprobe aus Heft 3-2014

In jeder Ausgabe bietet *Message* über zwanzig Beiträge zu aktuellen Fragen des deutschen und internationalen Journalismus. Immer sind einige dabei, die – durchaus gewollt – für Aufregung in den Redaktionen sorgen.

Um Ihnen einen kleinen Einblick in das Angebot der aktuellen Ausgabe zu geben, stellen wir immer einige Texte im Originallayout als PDF zum Herunterladen bereit. Dazu gehört der von Ihnen gewählte Beitrag.

Message erscheint vier Mal im Jahr. Zu jeder *Message* gehört immer auch das *Message*-Podium.

Sie können *Message* abonnieren: Im Internet unter der Adresse www.message-online.com oder per E-Mail unter message@evangemeindeblatt.de. Der Jahresbezugspreis beträgt 48,00 € (D/A/CH frei Haus, übriges Ausland plus Porto). Studenten/Volontäre und Mitglieder des »Netzwerk Recherche« zahlen nur 33,60 € (Nachweis erforderlich).

- *Message* beleuchtet die aktuellen Trends im Journalismus.
- *Message* bietet fundierte Analysen der deutschen und internationalen Medienwirklichkeit.
- *Message* ist praxisorientiert. Und selbstkritisch.

Im Jahr 2014 sind Roboter und Algorithmen im Journalismus nicht die Zukunft; sie sind die Gegenwart. Die Bilder einer Drohne gewannen in diesem Jahr einen World Press Photo Award; die iPhone-App »Wibbitz« formt auf Knopfdruck eine Online-Seite in einen Nachrichtenclip mit Bild und Ton um; und wenn man sie mit Daten füttert, erstellt die Software von »Narrative Science« Spielberichte und Unternehmensporträts, die sich tatsächlich so lesen, als wären sie von Menschen geschrieben. Wenn Programme das heute schon können – wie sieht dann erst die Zukunft des digitalen Journalismus aus?

Diese Frage trieb im März 2014 Programmierer, Wissenschaftler und Journalisten aus der ganzen Welt zur Konferenz »Groundbreaking Journalism« nach Berlin. Message präsentiert auf den folgenden Seiten eine Auswahl der Tagungsberichte, die uns die Veranstalter iRights.Lab und Vodafone Institut für Gesellschaft und Kommunikation freundlicherweise zur Verfügung stellten. Manches, was von den Autoren vorgezeichnet wird, mag man als diskussionswürdig ansehen, manches gar als bedrohlich. In abseh-

barer Zeit, schreibt der renommierte Datenjournalist Lorenz Matzat in seinem Beitrag (beginnend unten auf dieser Seite), werde Roboterjournalismus »für das, was heute im alltäglichen Nachrichtengeschehen als Journalismus gilt, allemal reichen«. Wer aber kontrolliert dann die Programme, die solche Texte schreiben? Auf der Podiumsdiskussion (ab Seite 65) wurde klar, dass es einer Clearingstelle bedürfte, einer Art moralisch-politischen TÜVs für journalistische Software. Welche Möglichkeiten sich derweil bieten, wenn Journalisten mit Informatikern zusammenarbeiten, zeigen Programme wie »Overview«. Mit dem Tool des amerikanischen Entwickler-Journalisten Jonathan Stray (s. Seite 67) lassen sich riesige Aktenmengen à la Wikileaks sortieren und strukturieren. Wer freilich glaubt, die Technik der jungen Generation überlassen zu können, wird auf Seite 69 widerlegt: Die Vorstellung vom »Digital Native«, der mit Handy und Tablet aufwächst und deshalb viel begabter für Technisches ist, ist nämlich falsch. Alle müssen sich also an Roboterjournalismus gewöhnen – nicht nur 1.0-Journalisten.

WAS IST EIGENTLICH ROBOTERJOURNALISMUS?

Noch sehen wir nur Ansätze, aber durch die technologische und ökonomische Entwicklung werden Algorithmen auch im Journalismus immer stärker verwendet. Sie könnten das Internet zumüllen und manipulieren – oder aber die journalistische Qualität heben.

Auch das noch: ein weiterer Journalismusbegriff. Nach Daten-, Drohnen- und Sensorjournalismus kommt jetzt auch noch Roboterjournalismus. Hierbei dreht es sich um Software, die in Teilbereichen des Journalismus selbstständig Artikel und Berichte erstellt. Vornehmlich wird er im Nachrichtenjournalismus, also der Berichterstattung stattfinden – und nicht bei Analyse, Kommentar, Interview und Investigation. Bislang

geschieht solcherlei nur experimentell, etwa bei *Forbes* (s. Artikel von Stefan Mey auf Seite 65). Doch bis 2020 dürften nicht wenige Jobs im Journalismus von Maschinen erledigt werden.

Warum dem so ist, will ich hier darstellen. Man mag sich streiten, ob Roboterjournalismus als Begriff taugt. Angesichts einer bislang mangelnden Alternative wird er hier für Algorithmen-Maschinen verwendet, die kompli-

zierte Arbeitsschritte zu einem gewissen Grad autark erledigen können.

ROBOTER ALS ASSISTENT

In der ersten Phase werden Roboter redaktionelle Assistenz leisten: Bei automatisierter Recherche, bei der Faktenüberprüfung, beim Scannen von Social Media und Sensornetzwerken; beim Mitlesen, -hören, -schauen von Websites, Radio- und TV-Sendern. Entspre-



Lorenz Matzat gehört zu den profiliertesten Entwicklern des Landes. Er gründete »Lokaler«, eine Firma, die Daten auf Karten abbildet.

Foto: David Jacob

Live-Ticker und Plattformen wie der Twitter-Sammeldienst Storify sein.

Nachrichtenagenturen liefern ihre Texte schon heute recht strukturiert aus, zum Beispiel als NewsML* (siehe Glossar). Diese Inhalte zuzuschneiden und in ein Redaktionssystem einzupassen, ist technisch einfach, weil inhaltlich gar nichts erfasst werden muss. Das allein wäre schlicht Automatenjournalismus. Doch auch die Spracherkennung wird immer besser – man schaue sich die automatisierte Untertitelung englischer Beiträge bei Youtube an. Politikerreden bei Phoenix, im Bundestag-TV oder andere Streams auszuwerten, kann Software übernehmen. Mittels semantischer Verfahren, dem Vergleich mit anderen Reden aus Archiven, aber auch durch Auswertung der Resonanz in sozialen Netzwerken wird Software die Kernaussagen von Reden identifizieren. Damit ließe sich der weitverbreitete »Verlautbarungsjournalismus« automatisieren, der mehr oder minder unhinterfragt Ausschnitte aus Pressekonferenzen wiedergibt.

Die zweite Phase im Roboterjournalismus wird dann eintreten, wenn die semantischen Fähigkeiten der Algorithmen so weit gediehen sind, dass sie in brauchbarer Qualität Beiträge für eine Vielzahl von Themenbereichen erzeugen können. Die kontrolliert dann noch ein Mensch vor Veröffentlichung und greift gegebenenfalls sprachlich etwas ein. Je mehr Informationen über Schnittstellen daher kommen, je mehr Vorgänge gemessen und beschrieben werden, umso mehr Bereiche der Berichterstattung werden sich automatisieren lassen. Und die Software wird immer besser werden, weil sie lernt. Es wird nicht zum vielgerühmten Qualitätsjournalis-

mus per Roboter kommen; aber für das, was heute im alltäglichen Nachrichtengeschehen als Journalismus gilt, wird es allemal reichen.

INHALTE ON THE FLY

Der Clou am Roboterjournalismus ist, dass er Inhalte *on the fly** erzeugen und ständig aktualisieren kann. Eine Individualisierung von Nachrichten wird so auf einem völlig anderen Niveau möglich. Nutzer erhalten anhand ihrer Schwerpunktsetzung auf ihre Hausnummer genau zugeschnittene Nachrichten. Ratsinformationssysteme ihrer Gemeinde, Pressemitteilungen, Agenturticker, Sportergebnisse, Terminkalender, Transkription von Radio, Fernseh- und Videobeiträgen, Wetter- und Verkehrssensoren in ihrer Umgebung und weitere Quellen werden dafür ausgewertet. Anhand dessen werden kurze Berichte verfasst und gegebenenfalls illustriert. Durch Sprachsynthese sind auch Podcasts möglich. Solche Audiostücke können wiederum automatisiert mit Bewegtbild, Fotos und Grafiken zu Videos zusammengeschnitten werden.

Bislang fehlen hierzulande noch die Softwareanbieter, die Roboterjournalisten zur Marktreife gebracht hätten. Im englischsprachigen Raum sind »Narrative Science« und »Automated Insights« in diesem Bereich unterwegs. Auch im deutschsprachigen Raum dürfte es bald Anbieter von solcher Software geben; die Technologie ist zumindest in Ansätzen vorhanden, etwa im Bereich der »Business Intelligence« oder bei Kanzleisystemen. Die Qualität der Roboterjournalisten-Software – der »engine«, die sich ein Medienanbieter leistet oder selbst entwickelt – wird ►

GLOSSAR

NEWSML

Datenformat zur Nachrichtenübertragung; strukturiert Elemente wie Text-, Foto-, Video- und Metainformationen und macht sie maschinenlesbar

ON THE FLY

simultane Erzeugung von Inhalten, zum Beispiel bei Javascript

SCRAPING

Daten sind im Internet häufig auf verschiedene Quellen verteilt oder lassen sich nur schwer herunterladen und verarbeiten. Beim »Scraping« werden sie mithilfe eines Programmes ausgelesen

SEMANTISCHES NETZ

eine Technik, mit der man Begriffe verknüpft darstellt; Form der digitalen Wissensrepräsentation

ein Kriterium dafür sein, welche Rolle er im Nachrichtenbereich spielen wird. Neben Anbietern von Technologieplattformen dürften sich hier auch neue Player etablieren, die keinen klassischen Verlag im Rücken haben. Ebenso können Medienmarken aus anderen Sprachräumen mittels Roboterjournalisten Sprachbarrieren einfacher überwinden.

HINTER DEM HYPE

Rein betriebswirtschaftlich wird sich das Ganze ab einem bestimmten Preis lohnen. Impuls könnte dafür hierzulande nicht zuletzt auch der Mindestlohn sein, weil sich in der ohnehin klammen Branche die Anschaffung von Roboterjournalisten-Software lohnt. Eine neue Stufe der Automatisierung durch Algorithmen und Roboter ist selbstredend nicht auf den Journalismus beschränkt; vielmehr wird dieser Wandel viele Branchen erfassen.

Manche werden das Gerede vom Roboterjournalismus als Technologieoptimismus oder -pessimismus abtun. Klar, Robotik ist ein Hype – spätestens seit das IT-Unternehmen Google in jüngster Zeit reihenweise Robotik-Unternehmen kauft. Doch Trends und Hypes sind nicht per se schlecht oder fabriziert; sie können versanden, können aber auch berechtigter Ausdruck des Zeitgeistes sein und manchmal auf einen epochalen Umbruch hindeuten.

Mehr und mehr integrieren sich Redaktion und IT; vor allem wenn das Medium rein digital erscheint. Entwickler und Programmierer erhalten Einzug in die Redaktionsräume – das wird auf Dauer die Kultur in Redaktionen ändern (s. Seite 44: Schwerpunkt zu journalistischen Kooperationen). Programmierer haben meist ein recht pragmatisches Verhältnis zu Informationen, und ihre Aufgabe ist es, Arbeit durch Softwaremaschinen erledigen zu lassen. Und wer je in einer Redaktion gearbeitet hat, dem dürften etliche Arbeitsschritte und Routineaufgaben einfallen, die sich dort noch automatisieren lassen könnten. Constanze Kurz und Frank Rieger schreiben im Epilog ihres Buches »Arbeitsfrei«: »Je weniger spezielle Talente und Fähigkeiten ein Arbeitsplatz erfordert, je besser sich Resultate messen, analysieren und quantifizieren lassen, desto direkter und unmittelbarer ist der Wettlauf mit den Maschinen«.

Apropos: »Race against the machine« hieß ein dünnes Büchlein, das Erik Brynjolfsson und Andrew McAfee vor gut zwei Jahren veröffentlichten. Ende 2013 haben die Ökonomen vom MIT ein weiteres Buch nachgelegt: »The Second Machine Age«. Man muss ihre doch sehr technologie-positivistische Sicht nicht teilen, die gesellschaftliche und soziale Fragen lediglich oberflächlich streift. Dennoch ist die Herleitung lesenswert, weshalb sie eine »zweite Industrialisierung« heraufdämmern

sehen. Ihrer Meinung nach ist die Geschwindigkeit, in der die Digitalisierung dank immer besserer und günstigerer Technologie voranschreitet, schneller als weitläufig angenommen. Die Kombination aus Forschung und Entwicklung in verwandten Gebieten sei im Begriff, durch ihre Synergie einen Technologieschub auszulösen. Vielleicht ist die Google-Street-View-Technologie, die – weil sie unscharfe Hausnummern erkennen muss – mittlerweile fast fehlerfrei Texterkennungs-Aufgaben (Captchas) löst, ein Indiz für diese These.

DAS INTERNET DER DINGE

Roboterjournalismus wäre uninteressant, wenn er nur dazu dienen könnte, Sportergebnisse zu verschriftlichen. Doch immer mehr Informationen liegen strukturiert vor oder werden in solche Form umgewandelt. Roboterjournalisten können direkt an das sagemaschinewebe »semantische Netz« andocken, mit dem man in der Informatik die Verbindung von Begriffen bezeichnet – siehe das Wikidata-Projekt.

Die unter Obhut des deutschen Wikimedia-Vereins entwickelte Datenbank sammelt Daten in maschinenlesbarer Form, sodass sie von anderen Systemen leicht weiterverwendet werden können. Wikidata wurde vor allem für die Wikipedia entwickelt. Dort kann die Datenbank beispielsweise dabei helfen, Angaben über die Einwohnerzahlen von Ländern zwischen verschiedenen Sprachversionen und Artikeln abzugleichen. Ein weiteres Beispiel ist das Vorhaben der BBC, ihre Inhalte granular, also sehr kleinteilig veröffentlichen zu können. Zudem wächst das »Internet der Dinge« heran – vielleicht viel langsamer und organischer, als die regelmäßigen Hype-Runden glauben machen wollen.

Unstrittig ist, dass die Zahl der Sensoren und mit dem Internet verbundenen Geräte stetig enorm steigt.

Roboterjournalisten können jede Regung und jeden Vorgang in dem entstehenden massiven Sensornetzwerk registrieren und gegebenenfalls beschreiben. Sie werden aus Bereichen berichten können, die Menschen gar nicht zugänglich sind. Vielleicht entstehen hier sogar neue journalistische Genres.

KAMPF DER ALGORITHMEN

Selbstredend birgt – wie jede Technologie – auch Roboterjournalismus Gefahren und kann missbraucht werden. Das große Vorbild »Narrative Science« wird von einem Venture-Capital-Arm des US-Geheimdienstes CIA mitfinanziert. Die deutsche Firma Aexea verdient ihr Geld unter anderem mit Suchmaschinenoptimierung (SEO) und dürfte ihre Technologie ursprünglich auch in diesem Bereich entwickelt haben. SEO ist letztlich ein spezieller Bereich der Werbewirtschaft, der mit lauterem oder unlauteren Mitteln die Position von Websites bei den Ergebnis-

sen von Suchmaschinen beeinflusst oder manipuliert. Hier herrscht ein ständiger Kampf der Algorithmen von Google und Co. mit denen der SEO-Fir-

In Zukunft könnte es eine Clearingstelle für Roboterjournalismus-Software geben.

men, die teilweise auf automatisch generierte Inhalte für »Linkfarmen« setzen oder auf halbautomatische Prozesse wie bei den Plattformen des Unternehmens »Demand Media«.

Das heißt: Die Methoden des Roboterjournalismus könnten beispielsweise eingesetzt werden, um eine automatisierte Schleife zum Fact-Checking für Beiträge vor Veröffentlichung zu haben; sie könnten helfen, die Qualität zu heben. Sie könnten andererseits auch dazu die-

nen, das Internet zuzumüllen oder gar systematisch und anhaltend Fakten zu verzerrern oder weitflächig zu manipulieren. Etwa durch koordiniertes und nimmermüdes Bearbeiten der Wikipedia auch über Sprachräume hinweg.

Neben den Auswirkungen der Roboter auf den Arbeitsmarkt nicht nur der Journalisten – denen sich Gewerkschaften und Politik beginnen sollten zu widmen – braucht es auch andere Vereinbarungen, die die Inhalte betreffen. Der Presse-Kodex könnte Maßstab für die Modellierung der Algorithmen sein, die unmittelbar für das Erstellen und Anreichern von Texten zuständig sind. Vielleicht gäbe es dann auch ein Zertifikat oder eine Clearingstelle für Roboterjournalismus-Software, die ihr ein journalistisches Verfahren bescheinigt, wenn nicht sogar einen Presseausweis ausstellt. ■

Lorenz Matzat

Dieser Beitrag erschien erstmals auf datenjournalist.de.

MEIN KOLLEGE, DIE MASCHINE

Schon heute können Programme Texte automatisiert schreiben. Andere Software kann helfen, Artikel vorzubereiten oder zu verbreiten. Doch wer kontrolliert diese Algorithmen?

Mitte März 2014 ereignete sich in Los Angeles eine kleine Sensation. Wenige Minuten nach einem Erdbeben in der Stadt erschien ein kurzer Artikel dazu auf der Website der Tageszeitung *Los Angeles Times*. Der stammte nicht wie üblich von einem Journalisten. Verfasst hatte ihn ein Algo-

rithmus, den ein Redakteur der Zeitung entworfen hatte. »Quakebot« hatte Geodaten analysiert. Als die auf ein Erdbeben hinwiesen, generierte er mithilfe von Textbausteinen eine kurze Nachricht. Der Redakteur war vom Erdbeben wach geworden, überflog den Text und veröffentlichte ihn dann.

Für Frederik Fischer ist das ein extremes Beispiel, wie Maschinen journalistische Aufgaben übernehmen können. Fischer ist selbst Journalist und Gründer von Tame. Der Dienst analysiert Twitter, um aus dem Rauschen der Tweets relevante Trends und Meinungen herauszufiltern. Fischer glaubt, dass Technologie den ►

journalistischen Prozess unterstützen kann: »Es gibt keinen Grund, weshalb wir Maschinen nicht für die Erledigung banaler Aufgaben nutzen sollten, sodass wir den Kopf für die interessanteren Dinge freibekommen.« Für verschiedene klassisch journalistische Aufgaben existieren bereits heute technologische Helferlein. »Talkwalker« etwa hilft bei der Beobachtung von Themen. Mit dem Dienst lassen sich Neuigkeiten zu individuell definierbaren Suchbegriffen verfolgen. Bei der Distribution von Inhalten hilft die App *Buffer*, die einen günstigen Zeitpunkt für die Veröffentlichung von Beiträgen in sozialen Netzwerken zu ermitteln verspricht.

Und dann gibt es auch noch ein Startup namens »Narrative Science«. Dessen Textgenerator lässt im großen Stil Artikel von Robotern schreiben. Dabei werden Ergebnisse von Footballspielen oder Unternehmensdaten in Textbausteine



Domenika Ahlrichs, Vize-Chefin von *Zeit Online*, im Gespräch mit Lorenz Matzat (re.), moderiert von Solana Larsen, Chefredakteurin von *Global Voices*.

eingefügt. Das Wirtschaftsmagazin *Forbes* lässt so beispielsweise Unternehmensporträts verfassen.

MASCHINEN BEREITEN VOR

Für Domenika Ahlrichs, stellvertretende Chefredakteurin von *Zeit Online*, wür-

den solche Einsatzszenarien zu weit gehen. Maschinen könnten Redaktionen zwar bei der Vorbereitung von Geschichten helfen. Dann aber müsse menschliche Intelligenz übernehmen.

Ein weiteres Szenario für die maschinell unterstützte Recherche sind datenjournalistische Storys. Als aktuelles Beispiel nennt Ahlrichs eine Visualisierung der ärztlichen Versorgung in Deutschland: »Wir haben Daten über spezialisierte Ärzte per Scraping* gesammelt und das dem Einkommen der Leute vor Ort gegenübergestellt. Aus dieser Kombination ergab sich eine interessante Geschichte über die ungleiche Verteilung von Ärzten in Deutschland.«

Dass Datensätze aber die Themenauswahl auf *Zeit Online* übernehmen, schließt Ahlrichs aus: »Wir nutzen natürlich Echtzeit-Statistiken über das Lese- und Klickverhalten bei Artikeln. Auf dieser Grundlage können wir leichte Anpassungen vornehmen, beispielsweise in Bezug auf Überschriften. Wir wollen aber immer noch selbst entscheiden, welcher Artikel gut für unsere Webseite

Fotos: David Jacob

ist.« Diesen eigenständigen Blick auf die Welt würde das Publikum von *Zeit Online* auch erwarten.

Im Jahr 2011 hatte der Datenjournalist Lorenz Matzat für *Zeit Online* eine Visualisierung zur Vorratsdatenspeicherung erstellt, die mit einem Grimme-Online-Award und dem Online Journalism Award ausgezeichnet wurde. Matzats aktuelles Projekt ist *Lokaler.de*, ein Softwareprojekt, das Daten auf Karten abbildet. Für die Berliner Stadtmagazine *Zitty* und *Tip* bereitet er deren Datenbank zu Veranstaltungen und Locations auf. Seine Erfahrungen mit Redaktionen fasst er folgendermaßen zusammen: »Programmierer, die in eine Redaktion kommen, bringen eine andere Kultur mit. Der Kern ihres Jobs

ist es, Prozesse zu automatisieren. Diese Automatisierung wird kommen. Das lässt sich nicht aufhalten.« (Siehe auch den Themenschwerpunkt Kooperation ab Seite 44.)

WER KONTROLLIERT?

In der Diskussion taucht immer wieder eine Frage auf: Wer kontrolliert dann eigentlich die Algorithmen, wenn sie eine so große Bedeutung im Journalismus erlangen? Wenn Algorithmen in Redaktionen Einzug halten und selbstständig Artikel schreiben, brauchen wir gut durchdachte Regeln für automatisierten Journalismus, meint Matzat. Die Frage nach der Kontrolle von Algorithmen stellt sich auch an einem anderen Punkt: Sollten Firmen ihre einfluss-

reichen Algorithmen offenlegen müssen, auch wenn das ihren Geschäftsinteressen widerspricht?

Frederik Fischer sieht hier vor allem ein Regulierungsproblem. Globale politische Strukturen fehlen, die große Firmen wie Google oder Facebook zu Transparenz zwingen könnten. Matzat hofft, dass bald eine Art Amnesty International diesen Job übernehmen könnte: »Wir brauchen eine NGO, die als Clearingstelle für die Kontrolle von Algorithmen fungieren kann.« Mit anderen Worten: Auch wenn Journalismus und Algorithmen zusammenwirken, sollte in der Partnerschaft der Mensch doch immer das letzte Wort haben. ■

Stefan Mey

DER JOURNALISMUS VON MORGEN

Das Projekt »Overview« von *Associated Press* soll helfen, Dokumentenberge auszuwerten. Software wie diese wird in Newsrooms bisher kaum verwendet. Entwickler-Journalisten wollen diese Kluft schließen.

Rund 5.000 Seiten, die ausgedruckt einen Papierstapel von knapp einem Meter Höhe ergeben würden – etwa drei Wochen müsste ein Journalist an den Berichten des US-Außenministeriums über den Einsatz von Sicherheitsfirmen im Irak lesen, um die Papiere durchzuarbeiten. Anhand solcher Dokumentenberge demonstriert Jonathan Stray, warum der Journalismus neue Software benötigt. Stray – Journa-

list, Informatiker und Fellow am Tow Center für digitalen Journalismus der Columbia-Universität – leitet das Projekt »Overview« von AP.

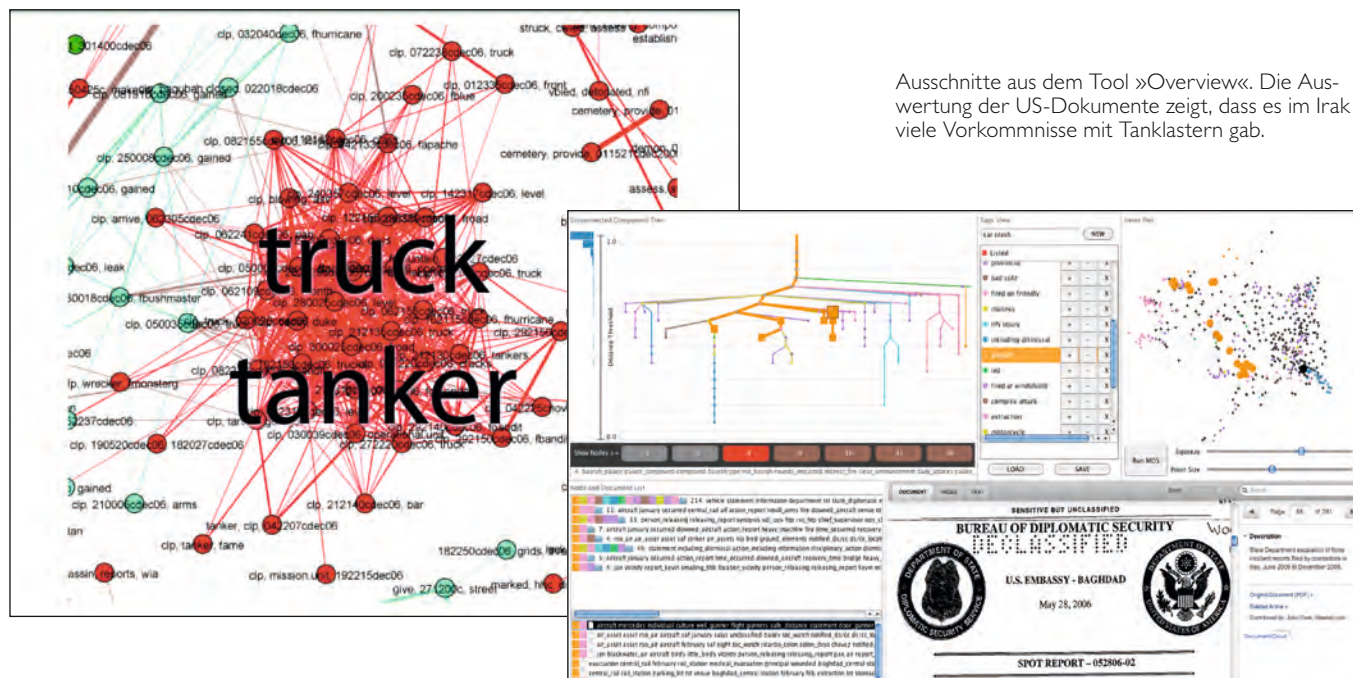
Füttert man das gleichnamige Programm mit einem solchen Stapel, untersucht es den Inhalt, sortiert die Dokumente und erzeugt eine Baumstruktur von Ordnern und Unterordnern anhand von Schlagworten. Hat es die Dokumente sortiert, kann der Nutzer sich von

Schlagwort zu Schlagwort hangeln und die Dokumente in Augenschein nehmen. Ein Ordner kann etwa das Schlagwort »Wahlen« tragen, der nächste »Detonation« – je nachdem, welche markanten Begriffe in den Dokumenten häufig vorkommen.

Die Idee für »Overview« entstand während der journalistischen Arbeit selbst: Stray durchforstete die von Wikileaks veröffentlichten »Iraq war ▶



Frederik Fischer gründete den Twitter-Aggregator *Tame* und ist Krautreporter für Netzkultur



Ausschnitte aus dem Tool »Overview«. Die Auswertung der US-Dokumente zeigt, dass es im Irak viele Vorkommnisse mit Tanklastern gab.

logs – 391.832 Militärdokumente zum Irakkrieg, ein noch größerer Berg als die Berichte des Außenministeriums. Wo sollte er anfangen?

Stray programmierte eine Visualisierung, die die Dokumente als Punkte zeigte – solche mit ähnlichen Inhalten verbanden sich zu Clustern. So fanden sich in den Dokumenten Muster, darunter einige hundert Erwähnungen von Vorfällen mit Tanklastzügen. Der Vergleich mit dem AP-Archiv ergab: Nur die Hälfte war dort dokumentiert. Stray drängte sich die Frage auf: »Hören wir nur von den Extremfällen – oder läuft es jeden Tag wie in den Meldungen?«

NUR SCHÖNE THEORIE?

Drei Wochen arbeitete Stray an der Visualisierung – ein Aufwand, den sich kaum eine Redaktion leisten kann. Eben das soll das »Overview«-Projekt abkürzen und Analyse-Werkzeuge anbieten.

Unter overviewproject.org steht das von der Knight Foundation unterstützte Projekt als kostenlose Webanwendung bereit. Dass hinter solcher Auswertungssoftware mathematische Modelle mit so klingenden Namen wie »multinomiale Dirichlet-Verteilung« oder die »inverse Dokumentfrequenz nach TF-IDF-Gewichtung« stecken, erwähnt Stray beiläufig. Entsprechende Algorithmen kommen in vielen Bereichen zum

Einsatz – bei Suchmaschinen oder Buch-Empfehlungen etwa. Für Stray gehören sie aber eben auch zum Werkzeugkasten des »Computational Journalism«, der nach seinem Verständnis vier Dinge leistet: rechnergestützt nach möglichen Geschichten schürfen, die Informationsflut bändigen, eine Geschichte durch Visualisierungen erzählen, die Verbreitung von Informationen nachvollziehen. »The revolution will be comma-delimited« steht auf dem Aufkleber seines Laptops; die Revolution wird kommagetrennt sein. Entwickler-Journalisten wie Stray lieben das Dateiformat »komma-getrennte Werte«, das Austausch und Analyse durch klar strukturierte Daten einfach macht und mit Excel ebenso wie mit Spezialanwendungen kompatibel ist. Doch solch »saubere« Daten sind eher selten als häufig anzutreffen. So fördert eine Informationsfreiheitsanfrage meist ausgedruckte und wieder eingescannte E-Mails zutage.



Screenshots: »Overview« (oben)
Foto: David Jacob (unten)

Jonathan Stray ist Informatiker und Journalist – eine Berufskoppelung, die leider selten ist.



Das Tool von Julian Kücklich: ein ganz analoger Stift. Auf der Groundbreaking-Journalism-Konferenz war Kücklich für das »Graphic Recording« zuständig.

Software-gestützter Journalismus: eine schöne, aber graue Theorie? »Es gibt eine riesige Lücke zwischen dem, was im Labor des Journalismus möglich ist, und dem, was im Newsroom auch umsetzbar ist«, räumt Stray ein. Auch die Technikgemeinde habe daran Anteil: Sie kümmere sich zu wenig darum, ihre schönen Entwicklungen auch praktisch einsatzfähig zu bekommen.

Software-gestützter Journalismus: eine schöne, aber graue Theorie? »Es gibt eine riesige Lücke zwischen dem, was im Labor des Journalismus möglich ist, und dem, was im Newsroom auch umsetzbar ist«, räumt Stray ein. Auch die Technikgemeinde habe daran Anteil: Sie kümmere sich zu wenig darum, ihre schönen Entwicklungen auch praktisch einsatzfähig zu bekommen.

KEINE DIGITAL NATIVES

Dass Software speziell für den journalistischen Einsatz entwickelt wird, ist ohnehin ein Sonderfall. Vorangetrieben wird die rechnergestützte Auswertung

nicht von Medienhäusern, sondern in anderen Bereichen. Zum Beispiel in der Medizin oder der Werbung, aber auch in den IT-Abteilungen von Geheimdiensten, wo die Datenberge am größten sind. Sind die Programme frei verwendbar und quelloffen, können sie für journalistische Zwecke angepasst und weiterentwickelt werden, merkt Stray an.

Aber nicht jeder Journalist müsse deshalb programmieren lernen, meint Michael Kreil, Entwickler bei der Datenagentur OpenDataCity. »Der beste Ansatz liegt in der Zusammenarbeit.« Kleine Teams von Journalisten und Entwicklern hält er für das beste Modell. Medienhäuser müssten sich daher nicht nur fragen: Welche Software brauchen wir? Sondern auch: Welche Software-Entwickler brauchen wir? Für Jonathan Stray ist das leichter gesagt als getan. Hochspezialisierte Datenarbeiter zu finden, die auch journalistisch versiert sind, gleiche »der Suche nach dem Einhorn.«

Wichtig sei ein permanenter Austausch zwischen Journalisten und Entwicklern, der über »Bau mir das!« hinausgeht, sagt auch Annette Leßmöllmann, Professorin für Wissenschaftskommuni-

kation am Karlsruher Institut für Technologie. Doch davon sei der Alltag im Newsroom noch weit entfernt. »Man muss erst einmal die Legende von den »Digital Natives« überwinden.« In der Journalistenausbildung an Hochschulen zeige sich: Nach der Einführung in journalistische Arbeitsweisen müsse man mit Grundlagen wie Twitter beginnen; erst Semester später sei dann noch Platz für den schwereren Stoff – wenn überhaupt.

Doch trotz aller neuen Techniken und Programme: Hergebrachte Arbeitsweisen werden durch sie nicht obsolet. Stray, der auch in Hongkong lehrt, erklärt seinen Studenten ebenso: Das Telefon ist eine Datenquelle. Nicht nur, aber gerade in China, wo deutlich weniger Daten offen zugänglich sind, muss man zum Hörer greifen. Auch die Verifizierung nehmen Programme wie »Overview« nicht ab. »Nur weil es Daten sind, sind sie noch nicht wahr. Da gelten immer noch die journalistischen Standards.« ■

David Pachali

Alle Texte unter Lizenz CC-BY-SA.

Links:

- ▶ www.forbes.com/sites/narrativescience/
- ▶ <http://narrativescience.com/>
- ▶ <http://automatedinsights.com/>
- ▶ <http://aexea.de>
- ▶ <http://overviewproject.org>
- ▶ www.bbc.co.uk/blogs/internet/posts/bbc-weather_getting_more_granu
- ▶ <http://slate.me/1fDbg56>