

Christoph Raetzsch

0 Structured Journalism

Unter dem Oberbegriff Structured Journalism wird eine Reihe von Anwendungen zusammengefasst, die eine Strukturierung und Beschreibung journalistisch gewonnener Informationen und Daten mithilfe von Metadaten vereinheitlichen sollen und so erweiterte Nutzungsformen hervorbringen können. Es handelt sich dabei weniger um ein neues Genre journalistischer Arbeit als vielmehr um eine sekundäre Dienstleistung, die der zunehmenden Digitalisierung journalistischer Produkte Rechnung trägt.

Da in Redaktionen fast nur noch datenbankbasierte Programme zur Erfassung, Darstellung und Planung journalistischer Erzeugnisse eingesetzt werden, bietet sich mit den Ansätzen des Structured Journalism eine Methode, separat voneinander abgelegte Informationen qualitativ zu beschreiben. Der Nutzen liegt dabei vorwiegend auf der Seite der Redaktionen und der Journalisten, da bereits vorhandene Informationen zu bestimmten Themen, Orten oder Personen einheitlich klassifiziert sind und so einfach abgerufen werden können. Im Rechercheprozess zu neuen Themen können Journalisten also auf bereits geleistete Recherchen ihrer Kollegen zurückgreifen.

Auch für Suchmaschinen werden veröffentlichte Informationen durch Metadaten qualitativ beschreibbar und, so die Hoffnung der Befürworter von Structured Journalism, erlauben eine zielgenauere Beantwortung von Suchanfragen der Nutzer, als dies mit der bis jetzt dominanten Suche nach Stichworten möglich ist.

Abgrenzung zum Datenjournalismus

Structured Journalism bezieht sich auf die Art, wie Informationen, die von Journalisten erhoben werden, nach möglichst einheitlichen Begriffen abgespeichert, strukturiert und wiederverwertet werden können. Dabei werden strukturierte Datensammlungen geschaffen – zu Personen,

Christoph Raetzsch

Institutionen und Ereignissen –, die ressort- und themenübergreifend als Ressource für verschiedene journalistische Darstellungs- und Produktformen genutzt werden können. Diese Herangehensweise ist nicht mit Datenjournalismus gleichzusetzen. Sie kann sogar als Gegenteil von Data Journalism oder Data-Driven Journalism (DDJ) bezeichnet werden.¹

Im Datenjournalismus werden typischerweise große Datenmengen („Big Data“) (halb-) automatisiert analysiert und ausgewertet. Diese mehr oder weniger einheitlich strukturierten Daten können aus offiziellen Statistiken stammen, wie sie im Zuge der Open-Data-Bewegung zunehmend auch öffentlich zur Verfügung gestellt werden. Sie können aber auch aus Dokumentenbeständen gewonnen werden, die von Whistleblowern im öffentlichen Interesse journalistischen Organisationen zugespielt werden (z. B. Wikileaks). Datenjournalismus oder Data-Driven Journalism wird betrieben, um über Auffälligkeiten bei der Datenanalyse und die Verknüpfung von Informationen aus geheimen oder bislang unbekanntem Quellen neue Zusammenhänge herzustellen und Stories zu generieren. So lassen sich neue Fragen aufwerfen und damit neue Themen für die weitere Recherche gewinnen.

Hintergrund

Der Begriff Structured Journalism wurde von Reg Chua geprägt, einem früheren Journalisten des Wall Street Journal. Er stellt in seinem Blog Structure of News² seit 2010 verschiedene Ansätze vor, Journalismus im Zeichen einer Absatz- und Aufmerksamkeitskrise in neue, attraktive und profitable Formen zu überführen. Vor allem regte Chua an, sich in der journalistischen Denk- und Arbeitsweise von der dominanten Form des Artikels zu distanzieren: „get past the story as the basic unit of journalism“³, forderte Chua damals.

In Artikelform seien die Produkte journalistischer Recherchen einem immensen Aktualitätsverfall preisgegeben, der sich über „copy-paste“-Verfahren von Text im Internet noch beschleunige. Mit zunehmender

¹ Siehe Beitrag „Datenjournalismus“ in diesem Band.

² <https://structureofnews.wordpress.com/>

³ Chua (2010).

Datenmenge und uneingeschränkter Verfügbarkeit von journalistischen Informationen über unzählige Plattformen tendiere der Wert eines Artikels gegen null, wenn er erst einmal veröffentlicht sei.

Suchmaschinen wie Google News stellen kostenfrei journalistische Produkte aus verschiedenen Quellen zur Verfügung, ohne dass die Aufmerksamkeit der Leser auch den Webseiten der Produzenten dieser Artikel zugute käme. Darüber hinaus müssen innerhalb der Redaktionen bereits recherchierte Informationen für eine spätere Verwendung immer wieder mühsam neu zusammengesucht werden, denn in der narrativen Form des Artikels lassen sich einzelne Informationen nicht in einheitlicher Form ablegen oder über Suchbefehle herausfinden: Ist eine konkrete Information erst einmal in einem Satz aufgeschrieben, kann sie nicht mehr automatisch wiedergefunden werden.

Chua forderte daher: „Structured Journalism is to change the way we create content so as to maximize its shelf-life, as well as structuring – as much as possible – the information in stories, at the time of creation, for use in databases that can form the basis of new stories or information products.“⁴

Ansätze des Structured Journalism

Im Redaktionsbetrieb von Magazinen, Zeitungen und Fernsehanstalten finden datenbankbasierte Programme zur Planung, Herstellung und Archivierung von Nachrichten heute vielfach Verwendung. Über Redaktionsdatenbanken werden Inhalte erstellt, in Content-Management-Systemen (CMS) werden diese Inhalte im Internet veröffentlicht.

Allerdings können CMS, als Schnittstelle zwischen dem Inhalt und seiner Darstellung im Internet, diese Inhalte nicht qualitativ unterscheiden; sie strukturieren nur die Darstellung bestimmter Artikel und Webseitenelemente wie Überschriften, Autorennamen, Textblock, Datum der Veröffentlichung. Die Beziehungen zwischen grafisch aufbereiteten Inhalten (Überschrift, Textblock, Autor usw.) müssen qualitativ in Metadaten beschrieben sein, um eine je andere Unterscheidung von wichtigen und unwichtigen Elementen auch auf Seite der Suchmaschinen und -algorithmen abbilden zu können.

⁴ Chua (o.D).

Christoph Raetzsch

Jeder Leser und jede Zuschauerin beurteilt angebotene Informationen nach eigenen (und oft auch unvorhersehbaren) Kriterien von Relevanz. Ein CMS dagegen stellt alle Informationen auf einer Webseite gleichwertig dar, was dazu führt, dass zum Beispiel Suchmaschinen alle Daten einer Webseite aggregieren, also zusammenstellen, statt sich auf Namen, Orte, Ereignisse, also für Leser sinnvolle Informationen, zu konzentrieren.

Für die Nutzer von journalistischen Angeboten geht es aber um die Bedeutung von Informationen, nicht primär um deren grafische Darstellungsweise. Um diese Bedeutung, zumindest ansatzweise, auch für eine Suchmaschine oder einen Algorithmus verfügbar zu machen, müssen einzelne Informationen mit eindeutigen Metadaten, das heißt Klassifikationen ihres Inhalts, versehen werden.

Hier setzt Structured Journalism an, indem er die Strukturierung journalistischer Informationen im Recherche- und Veröffentlichungsprozess und die Ebene der qualitativen Metadaten zusammenbringt, um journalistische Erzeugnisse über metadatenbasierte Anwendungen für andere Anbieter verfügbar zu machen.

Structured Journalism und Linked Data

Structured Journalism stellt damit eine spezielle Anwendung der Prinzipien von *Linked Data* und des *Semantic Web* dar, die eine Beschreibung von Daten und deren Verknüpfung über verschiedene digitale Anwendungen und Online-Plattformen hinweg erreichen wollen.

Diese Strukturierung erfolgt durch Metadaten, die es erlauben, innerhalb des dominanten HTML-Standards (Hypertext Markup Language) zur Darstellung von Webseiten Aussagen über die Klasse, Kategorie oder Art von Informationen einzubetten. Eine solche Klassifizierung von Daten ist die Grundlage von Anwendungen mit Linked Data, da so plattformübergreifend Daten neu zusammengestellt und analysiert werden können.

Allein auf Basis des HTML-Standards lassen sich solche strukturierten Datensammlungen aber nicht erzeugen.⁵ Linked Data bezeichnet die konkrete Verwendung semantischer Verfahren zur Klassifizierung von Inhalten natürlicher Sprachen, wie sie im Zusammenhang mit dem Semantic Web erforscht werden.

Der Ansatz des Semantic Web „allows computers to intelligently search, combine, and process Web content based on the meaning this content has to humans“⁶. Es geht also darum, Hierarchien, Abhängigkeiten und Beziehungen zwischen Worten natürlicher Sprache, von Begriffen und Akteuren so zu beschreiben, dass sie mittels computergestützter Datenverarbeitung als solche erfasst und mit logischen Operationen verarbeitet werden können.

Für einen Computer oder Suchalgorithmus unterscheidet sich eine Reihe von Buchstaben (auch *string* genannt) wie „Merkel“ und „merken“ nur minimal. In ihrer Bedeutung aber befinden sich beide Worte semantisch und grammatikalisch auf verschiedenen Ebenen. Auf Computerseite ist eine qualitative Unterscheidung nur durch eine strukturierte Klassifizierung von Attributen wie „Eigennamen“, „Verb“ oder „Staatsoberhaupt“ zu erreichen. Auf Grundlage einer Klassifizierung lassen sich dann Gruppen ähnlicher Begriffe (Eigennamen oder Verben) bilden und im Anschluss wiederum Querschnittsmengen oder Beziehungen herstellen, zum Beispiel alle Eigennamen, die auch Staatsoberhäupter sind.

Der Ansatz des Structured Journalism als Anwendung von Linked-Data-Prinzipien erfordert dabei zunächst, Metadaten vorab festzulegen und in einem Klassifizierungsschema (einer sogenannten Ontologie oder Taxonomie) zu gliedern. Diese Schemata ordnen relevante Informationen von allgemeinen hin zu speziellen Ausprägungen, ähnlich einem Schlagwortkatalog, der zur Klassifizierung von Forschungsliteratur in Bibliotheken eingesetzt wird.

Im Journalismus werden derzeit vor allem zwei Schemata zur Bildung von Metadaten verwendet: rNews und NewsML-G2, das Austauschformat des

⁵ Heath und Bizer (2011), S. 2.

⁶ Hitzler, Kröttsch und Rudolph (2009), S. 11.

Christoph Raetzsch

International Press Telecommunications Council (IPTC)⁷. Dabei basiert das rNews-Schema auf dem allgemeineren RDFa-Schema (Resource Description Framework in Attribute) des W3-Council, einem internationalen Konsortium für Standards im Internet;⁸ rNews wurde speziell auf die Erfordernisse von Nachrichtenorganisationen hin als übergreifende Klassifizierung von Metadaten entwickelt.

Anwendungen

In einem Pilotprojekt bei der BBC dienten RDFa-Metadaten vor allem der Optimierung von Suchanfragen, die im Quellcode der Webseite erfolgte und für den Nutzer zunächst relativ wenig Bedeutung hat: „this additional code does not directly impact the user's experience of the page, [but] it makes a big difference to machines,“ schreibt Jeremy Tarling, Datenarchitekt in dem Projekt.⁹ Für die Nutzer bleibt zunächst alles beim Alten. Allerdings lassen sich aufgrund durchgängiger Verschlagwortung relevanter Informationen in Artikeln und Beiträgen neue Anwendungen generieren, die qualitativ höheren Nutzen haben als eine Suchanfrage, die allein auf strings oder Buchstabenketten beruht. Über Suchoperatoren und die Einbindung der Metadaten in Suchanfragen ließen sich direkt Artikel „über“ ein Thema von Artikeln unterscheiden, in denen nur an irgendeiner Stelle der „Begriff“ oder das Thema genannt sind.¹⁰

Eine der Hoffnungen hinter Structured Journalism ist sicher, jenseits der Datenmacht großer Suchmaschinen wie Google eigene Produkte zu gestalten, die auf der analytischen und sachlichen Kompetenz von Journalisten beruhen, und diese in übersichtlicher und relevanter Form den Nutzern zu präsentieren.

⁷ http://www.iptc.org/site/News_Exchange_Formats/NewsML-G2/Specification/;
<http://dev.iptc.org/rNews>; <http://dev.iptc.org/Introduction-To-RDF>

⁸ <http://www.w3.org/TR/xhtml-rdfa-primer/>

⁹ Tarling (2013).

¹⁰ Tarling (2014).

Besonders in großen Nachrichtenorganisationen wie der BBC oder der New York Times¹¹ ist es unmöglich, alle Inhalte aller Plattformen, Kanäle und Distributionsformen parallel im Blick zu behalten. Über die strukturierte Einbettung von Metadaten wird eine Verknüpfung dieser Inhalte, so die Hoffnung, nun möglich und unter Umständen Grundlage neuer journalistischer Produkte. So schreibt Oliver Bartlett, Produktmanager bei der BBC: „This means the [linked data] platform doesn't just know that tomorrow's episode of the *Culture Show* features *Jarvis Cocker*. It also knows that *Jarvis* is from *Sheffield*, was the lead singer in *Pulp*, that *Pulp* were a *Britpop* band, that they had a single called *Common People*, and that *Common People* was played on *6 Music* this morning. We can suddenly start to make connections across our products and content which couldn't previously have been made without a lot of manual effort.“¹² Für jedes Stichwort (hier: kursiv) lassen sich Verknüpfungen innerhalb aller verfügbaren Angebote der BBC herstellen, was sowohl auf Journalisten- wie auf Nutzerseite zu einem breiteren Angebot an Informationen führt, die innerhalb kurzer Zeit verfügbar gemacht werden können.

Die konsequente Einbettung von Metadaten kann zum Beispiel in der Automatisierung der Nachrichtenproduktion genutzt werden, gerade dort, wo es aufgrund der überschaubaren Anzahl an Ereignissen wenig Variation auf sprachlicher Seite gibt. In Wetternachrichten, in Berichten über Sportereignisse oder den Aktienmarkt tauch meist eine begrenzte Anzahl stets wiederkehrender Elemente auf. Oft basieren diese Arten von Nachrichten auch vorwiegend auf Zahlen oder anderen quantifizierten Informationen. Unternehmen wie Narrative Science bieten zur automatischen Generierung von Artikeln bereits Software an, die auf Basis von Rohdaten und Sprachalgorithmen gut lesbare Artikel verfasst. Dies wird im amerikanischen Big Ten Network, einem Kabel-TV-Netzwerk für Sportübertragungen, bereits seit 2010 für Spielzusammenfassungen getestet.¹³

¹¹ Siehe auch die Präsentation von Evan Sandhaus, dem leitenden Entwickler semantischer Plattformen bei der New York Times, zur Einbindung von rNews: <http://www.slideshare.net/sandhaus/rnews-0-7abridgedexpanded>.

¹² Bartlett (2013).

¹³ Lohr (2011).

Christoph Raetzsch

Am häufigsten finden sich bislang Anwendungen von Structured Journalism bei der Verlinkung von Artikeln über Schlagworte innerhalb der Webseite eines Anbieters. Bei neuen Ereignissen lassen sich schnell Themenseiten generieren, auf denen bereits geschaffene Inhalte aktiv gehalten und weitergehend verwertet werden können.

Ferner bietet sich durch Liveblogs zu aktuellen Ereignissen die Möglichkeit, in kurzen Abständen neue Mikroeinheiten mit aktuellen Informationen zu veröffentlichen, die auch Reaktionen und Diskussionen in sozialen Medien integrieren.¹⁴ Andere Anwendungen¹⁵ bieten sich als Plattform an, um eine Vielzahl von an sich disparaten Informationen zu sammeln, zeitlich zu strukturieren und in größere Erzählstränge einzubinden. Verschiedene Datenquellen, die Journalisten in ihrer Arbeit verwenden, können so gesammelt und strukturiert abgebildet werden, um eine neue Grundlage journalistischer Produktion zu werden.

Wenn ein Ereignis ein Nachrichteninteresse auslöst, können Daten aus vielen Quellen schnell bereit gestellt werden. Zu einem übergeordneten Thema lassen sich verschiedene Erzählstränge bilden, die ihrerseits zur Erhebung neuer Daten führen.¹⁶

Mit dem Ansatz des Structured Journalism lässt sich eine Vielzahl von Informationen neu verknüpfen, die die journalistische Arbeit selbst vereinfachen und auch ihre Verwertung in neuartigen Nachrichtenprodukten als Dienstleistung abseits etablierter Publikationsformen möglich machen.

Vorteile und Nachteile von Structured Journalism

Unter dem Oberbegriff Structured Journalism werden gegenwärtig verschiedene Ansätze zur datenbasierten Innovation journalistischer Produkte diskutiert. Dazu existiert eine Reihe von Pilotversuchen, vor allem

¹⁴ Beispielsweise auf verge.com oder vox.com.

¹⁵ Wie z. B. storify.com.

¹⁶ Steglich, Nagy und Horn (2014); Oswald (2014).

großer Nachrichtenorganisationen, die die Einbettung von Metadaten und deren Analyse in der journalistischen Arbeit praktisch umsetzen sollen.

Zum gegenwärtigen Zeitpunkt lassen sich zu diesen fortschreitenden Entwicklungen im Journalismus noch keine Bewertungen abgeben. Dennoch erscheint das Ziel klar, dass mithilfe von strukturierten Daten Nachrichten und andere journalistische Produkte besser und profitabler mit anderen Webanwendungen verknüpft werden können. Beziehungen zwischen Artikeln oder Beiträgen lassen sich intern und extern leichter finden; individuelle und oft nur cursorische Informationen können über Metadaten in strukturierter Form abgelegt und wiedergefunden werden. Neue Zusammenhänge zwischen unabhängig voneinander recherchierten Geschichten lassen sich mit erhöhter Komplexität der Datenbank einer Redaktion leichter herstellen.

Die Synergien sowohl im Redaktionsgeschäft als auch im Hinblick auf die Verwertung journalistischer Informationen, die Structured Journalism verspricht, basieren aber auf einer erheblichen Veränderung journalistischer Praktiken und Routinen.

Zunächst ist der Aufwand zu Beginn einer Datensammlung bzw. die Umstellung eines Redaktionssystems auf die strukturierte Erhebung von Metadaten aufwendig und setzt signifikante Entwicklungsinvestitionen voraus. Darüber hinaus ist bislang unklar, ob Structured Journalism eine Form von Zusatzqualifikation von Journalisten oder ob für die Implementierung und Pflege von Metadaten neue Berufsbilder an der Schnittstelle von Software- und Datenbankentwicklung nötig sind, die auf die Anforderungen journalistischer Arbeits- und Darstellungsweisen spezialisiert sind.

Vor dem Hintergrund der pluralisierten Informations- und Kommunikationsumgebung des Internets und sozialer Medien im Besonderen erscheint es sicher sinnvoll, die Dominanz des Artikels als Maßstab und Norm journalistischer Arbeit zu überdenken. Allerdings knüpft sich an diese Darstellungsform auch eine kulturelle Wertung von journalistischer Autorität, die nicht allein durch professionelle Umstrukturierung sogenannter Inhalte wie Informationen und Daten zu ersetzen ist. Die meisten Praktiker im Journalismus werden Informationen über gesellschaftliche Prozesse und Akteure nicht als Daten auffassen,

Christoph Raetzsch

sondern als Wissen über die Gesellschaft, in der sie leben und die sie kritisch und politisch beobachten.

Mit Structured Journalism soll ausgelotet werden, wie Journalisten ihre vielfältigen Expertisen und Informationen auf neuen technologischen Wegen der Öffentlichkeit zur Verfügung stellen können. Ob es sich bei diesem Ansatz um eine technische oder gesellschaftliche Innovation handelt, kann erst die Entwicklung neuer journalistischer Formen zeigen.

Quellen

- Bartlett, O. (2013): Linked Data: Connecting together the BBC's Online Content. BBC Internet Blog. 19.02.2013: <http://www.bbc.co.uk/blogs/internet/entries/af6b613e-6935-3165-93ca-9319e1887858>.
- Chua, R. (2010): Getting Past the Story. (Re)Structuring Journalism. 08.08.2010: <https://structureofnews.wordpress.com/2010/08/08/getting-past-the-story/>
- Chua, R. (o.D): "Structured Journalism". <https://structureofnews.wordpress.com/structured-journalism/>.
- Gillespie, T (2014): The Relevance of Algorithms, in: Gillespie, T; Boczkowski, P; Foot, K. (Hrsg.). Media Technologies. Essays on Communication, Materiality, and Society, 167-194. Cambridge, MA..
- Heath, T; Bizer, C. (2011): Linked Data. Evolving The Web into A Global Data Space, San Rafael. <http://linkeddatabook.com/editions/1.0/>.
- Hitzler, P; Krötzsch, M; Rudolph, S. (2009): Foundations of Semantic Web Technologies, Boca Raton.
- Lohr, S. (2011): In Case You Wondered, a Real Human Wrote this Column, in: New York Times, 10.09.2011: <http://www.nytimes.com/2011/09/11/business/computer-generated-articles-are-gaining-traction.html>.
- Oswald, B. (2014): Structured Journalism: Journalisten als Bibliothekare. <http://get.torial.com/blog/2014/05/structured-journalism-journalisten-als-bibliothekare/>.
- Steglich, F; Nagy, T; Horn, S. (2014): Assets, Objects, Points: Was Structured Journalism Bringen Kann. Vortrag auf der re:publica, 07.05.2014: <http://bit.ly/17V4C2q>.
- Tarling, J. (2014): Storylines as Data in BBC News. Medium.com. 01.04.2014. <https://medium.com/@jeremytarling/storylines-as-data-in-bbc-news-bd92c25cba6b>.

Tarling, J. (2013): BBC News Linked Data Ontology. bbc.co.uk. 03.06.2013.
<http://www.bbc.co.uk/blogs/internet/entries/0e8d8442-defc-39e4-85bf-e04a385c02c9>.