arbido

2020/ Abbild und Inszenierung der Gesellschaft

Schubert Irina, Archivinformatikerin bei Docuteam

Tags: Archiv Social Media, Schlaglicht,

Wie und weshalb wir Twitter-Timelines archivieren

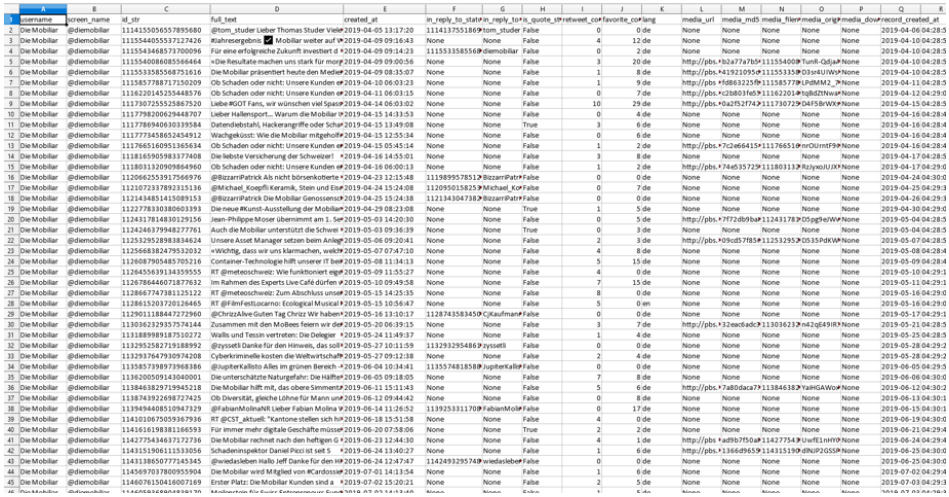
Twitter verfügt über grosses historisches Potenzial. Selbst wenn auf der Plattform sämtliche Tweets bis in die Anfangszeit zurück abrufbar sind, ist die Verfügbarkeit für die längerfristige Zukunft nicht gesichert. Bei docuteam haben wir ein Verfahren entwickelt, Twitter-Timelines zu archivieren.

Fast jeder kennt Twitter, die Microblogging-Plattform, auf der Informationen in Realtime verbreitet werden, Communities und Meinungen gebildet – und ebenso manipuliert werden. Die Publikationsgeschwindigkeit ist hoch, die Reaktionszeit kurz. Das macht Twitter zu einem Brandbeschleuniger für Debatten aller Art. Twitter erlaubt eine Demokratisierung der Berichterstattung, ermöglicht Meinungsfreiheit und dient der Mobilisierung politischer und gesellschaftlicher Bewegungen. Es sorgt aber auch für die Verbreitung von Fake-News, Hasskommentaren und Belästigungen.

Der erste Tweet: "just setting up my twttr" wurde vom Mitbegründer und heutigem CEO Jack Dorsey am 21. März 2006 abgesetzt. "twttr", so hiess der Prototyp damals, wurde eigentlich als Kommunikationsmittel für ein anderes Projekt entwickelt. [1] So zumindest eine populäre Legende. Die Gründer öffneten ihre Plattform bald einer breiten Nutzerschaft – und waren über den Erfolg selbst überrascht. [2] Twitter Inc. wurde ein Jahr später gegründet und entwickelt das Produkt mit mittlerweile rund 5000 Mitarbeitenden noch heute weiter. Bereits 2007 wurden Hashtags eingeführt, um Themen zu strukturieren. Als SMS-kompatibler Dienst entwickelt, waren lange nur 140 Zeichen erlaubt (die restlichen 20 waren für den Usernamen reserviert). Seit 2017 sind Tweets von einer Länge bis 280 Zeichen möglich. Heute (Stand 1. Quartal 2020) zählt der Dienst 166 Mio. sogenannte mDAU («monetizable Daily Active Usage»). [3] Eine Zahl, die seit Jahren stetig wächst. Monatlich werden über 500 Mio. Nutzer*innen gezählt, die Twitter besuchen, ohne sich einzuloggen. [4] Gewinn macht Twitter Inc. erst seit Ende 2017 und zwar vor allem über Werbung. [5][6] Der Gesamtumsatz im 1. Quartal 2020 liegt bei über 800 Mio. US-Dollar. [7] Pro Sekunde werden aktuell durchschnittlich etwa 6000 Tweets abgesetzt. Das macht rund 500 Mio. am Tag. Die meisten Follower hat aktuell Barack Obama.

Twitter ist auch ein Ort der Selbstinszenierung. Dies gilt für Einzelpersonen ebenso wie für Unternehmen und öffentliche Einrichtungen. Es werden Beziehungen gepflegt und Marketing betrieben. Für das konstante Erstellen und Einspeisen von Inhalten wird viel Geld investiert. Die Kommunikation ist weniger formell als in gedruckten Medien und geht zudem in zwei Richtungen: Posts werden nicht nur gelesen, sondern auch zeitnah kommentiert, «geliked» und «retweetet».

Was bringt unsere Kund*innen dazu, ihre Twitterkanäle archivieren zu lassen? Neben der empfundenen historischen Relevanz komplettiert das Social-Media-Archiv die klassischen Bestände gedruckter Kommunikationsmittel. Ohne den digitalen Kanal wäre das Bild der Kommunikationstätigkeit unvollständig. Daneben gibt es auch ganz praktische Argumente. Beispielsweise ist es für ein Unternehmen wichtig, den genauen Zeitpunkt und Wortlaut einer öffentlich kommunizierten Mitteilung nachvollziehen und belegen zu können. Sind die Twitterdaten erst einmal archiviert, haben sie eine Prüfsumme. Die Daten können so im Nachhinein nicht unbemerkt verändert werden.

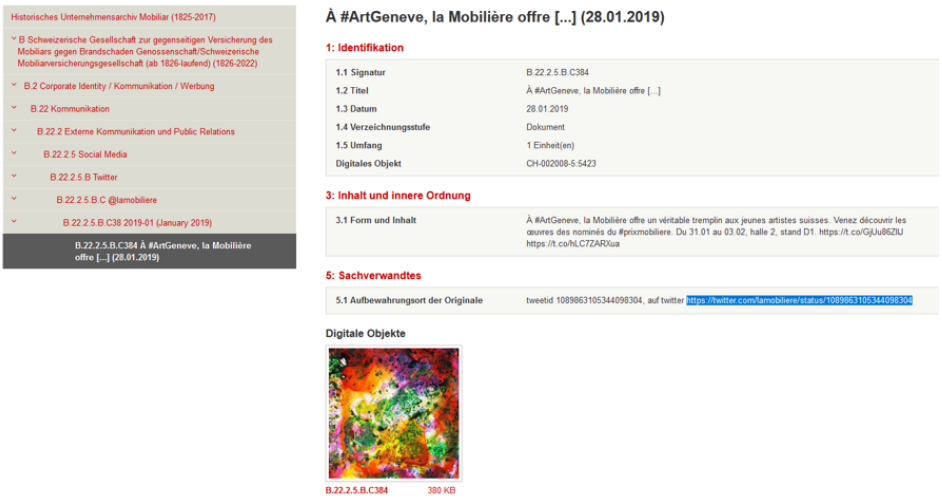


Metadata

Seit dem ersten Tweet sind 14 Jahre vergangen. Da die Plattform ihre Daten bis in die Anfangszeit zurück zugänglich macht (ausser die Kontoinhaber*in löscht diese selbst), wird Twitter oft auch als Archiv bezeichnet. Dass man darauf mit Blick auf die Langzeitarchivierung nicht zählen sollte, muss hier nicht weiter begründet werden. Was zeigt uns ein Blick in unsere Branche? Das bekannteste Projekt war vermutlich die Archivierung sämtlicher Tweets durch die Library of Congress im Jahr 2010. Es stiess auf grosse Begeisterung, obwohl nur Text archiviert wurde, löste aber auch einige Skepsis in Bezug auf Datenschutz aus. So stehen die archivierten Tweets in ihrer Gesamtheit bis heute unter Verschluss. Das Projekt fand sieben Jahre später sein Ende. Seither archiviert die LoC nur noch ausgewählte Tweets.

Was bedeutet das für Archivar*innen? Es bedeutet, dass wir uns selbst um die Archivierung kümmern sollten, wenn uns unsere Tweets lieb (das heisst: archivwürdig) sind. Twitter bietet mehrere Programmierschnittstellen an, mit denen sich die Daten automatisiert abgreifen lassen.[8] Daneben gibt es Unternehmen, die eine detaillierte Auswertung von Twitter- und anderen Social-Media-Daten beispielsweise für Marketing- oder politische Zwecke anbieten. Dafür zahlt man aber sehr viel Geld. Und sind die Daten damit wirklich gesichert? Bedingt.

Was machen wir bei docuteam? Wir speichern die Tweets mitsamt Metadaten in eine Datenbank und überführen diese Daten periodisch in ein digitales Langzeitarchiv. Zusätzlich können die heruntergeladenen Daten über eine einfache passwortgeschützte Web-Oberfläche jederzeit angesehen und durchsucht werden. Die letzten 3200 Tweets erfahren täglich ein Update, was die Anzahl Retweets und Likes anbelangt. Kommentare werden nicht archiviert. Neben dem Text des Tweets wird zusätzlich jeweils ein Bild, sofern vorhanden, mitgespeichert. Das Herunterladen von Videos lässt Twitter zurzeit nicht zu. Für das Abgreifen der Daten benutzen wir die Twitter API, für die Archivierung in der Regel unsere eigenen Tools.



Post Katalog

Sobald die Daten ins Langzeitarchiv überführt worden sind, sind sie fixiert. Das heisst, sie erfahren keine Aktualisierung mehr, befinden sich in der sicheren Umgebung eines digitalen Langzeitarchivs und sind über das Archivinformationssystem auffindbar.

Unsere bisherige Erfahrung zeigt, dass unser Ansatz für die Twitterarchivierung besonders bei Unternehmensarchiven auf Interesse stösst. Wir sind aber der Meinung, dass die Archivierung von Twitter- und allgemein Social-Media-Daten grundsätzlich für jedes Archiv eine Überlegung wert ist.

[1] Reissmann, Ole et al.: «Meilensteine des Weitererzähl-Webs» [online], 2011, https://www.spiegel.de/netzwel... (Zugriff am 19.05.2020)

[2] Ebd.

[3] Twitter Inc.: «Quarterly results» [online], 2020, https://investor.twitterinc.co... (Zugriff am 19.05.2020)

[4] Rao, Deepak: «Testing promoted Tweets on our logged-out experience» [online], 2015, https://blog.twitter.com/en_us... (Zugriff am 19.05.2020)

[5] Kollewe, Julia: «Twitter makes first quarterly profit in its history» [online], 2018, https://www.theguardian.com/te... (Zugriff am 19.05.2020)

[6] Gadkari, Pia: «How does Twitter make money?» [online], 2013, https://www.bbc.com/news/busin... (Zugriff am 19.05.2020)

[7] Twitter Inc.: «Quarterly results» [online], 2020, https://investor.twitterinc.co... (Zugriff am 19.05.2020)

[8] Twitter Developer Documentation [online], 2020, https://developer.twitter.com/... (Zugriff am 19.05.2020)

Irina Schubert

Archivinformatikerin bei Docuteam GmbH.

Abstract

Deutsch

Twitter verfügt über grosses historisches Potenzial. Selbst wenn auf der Plattform sämtliche Tweets bis in die Anfangszeit zurück durchsuch- und abrufbar sind, ist die Verfügbarkeit für die längerfristige Zukunft nicht gesichert. Das wissen wir in unserer Branche nur zu gut. Bei docuteam haben wir deshalb ein Verfahren entwickelt, Twitter-Timelines für Archive, die von uns betreut werden, zu archivieren.

Français

Twitter a un grand potentiel historique. Même si tous les tweets de la plateforme peuvent être recherchés et récupérés depuis les premiers jours, leur disponibilité à long terme n'est pas garantie. Dans notre secteur, nous ne le savons que trop bien. A docuteam, nous avons donc développé une procédure d'archivage des calendriers Twitter pour les archives que nous gérons.