

Nachhaltige Archivierung, Erschließung,
Bereitstellung dynamischer Daten aus
sozialen Medien - Twitter und danach

19. - 20. März 2024

Deutsche Nationalbibliothek

Dr. Peter Worm, Stadtarchiv Münster
Marc Malwitz, LWL.IT



Erfolge und Grenzen bei der Archivierung von Twitter-Timelines

Ein Erfahrungsbericht aus
kommunaler Perspektive

2019

Vorüberlegungen und erste Schritte

Preliminary considerations and first steps

Archival material or trash?!

Ist das Archivgut oder kann das weg?!

← **Markus Lewe**
683 Tweets




Markus Lewe
@LeweMarkus

Oberbürgermeister der Stadt Münster. Impressum: t1p.de/pklt

📍 Münster, Deutschland stadt-muenster.de/rathaus/oberbu...
📅 Seit Januar 2012 bei Twitter

230 Folge ich 3.966 Follower

 Gefolgt von Smart City Münster, Junges Forum DIG Münster und 18 weiteren Personen, denen du folgst

Tweets Tweets und Antworten Medien Gefällt mir

/ml

???

/tl

 **Markus Lewe** @LeweMarkus · 14. Feb. 2019 ...
Ziel ist eine gute Lösung für das Hafen-/Hansaquartier. Die Bürgerschaft erhält im "Heilungsverfahren" die Möglichkeit, sich mit Stellungnahmen einzubringen. /ml

 **Stadt Münster** @muenster_de · 13. Feb. 2019
Neues Beteiligungsverfahren für Bebauungsplan Hansa/Hafen voraussichtlich schon ab Mitte 2019 - alle Infos in der Pressemitteilung: muenster.de/stadt/pressese...

 **Markus Lewe** @LeweMarkus · 30. Aug. 2020 ...
Ein unverhofftes Geschenk der Geschichte, dass der Reichstag zu einem Zentrum von Freiheit und Demokratie wurde. Umsomehr erschüttern mich die Bilder, wie Rechtsextreme versuchen, den freiheitlichen Geist dieses Ortes anzugreifen. Jetzt erst recht Frieden und Freiheit stiften!

6 8 109

 **Markus Lewe** @LeweMarkus · 17. Apr. 2020 ...
In unregelmäßigen Abständen informiert Oberbürgermeister Markus Lewe über die Arbeit im Krisenstab und die Entwicklung des Coronavirus' in der Stadt Münster.
Hier das Update vom 17.04.2020. /tl
youtu.be/suQenj6p7B4

Are we allowed to do all of this?

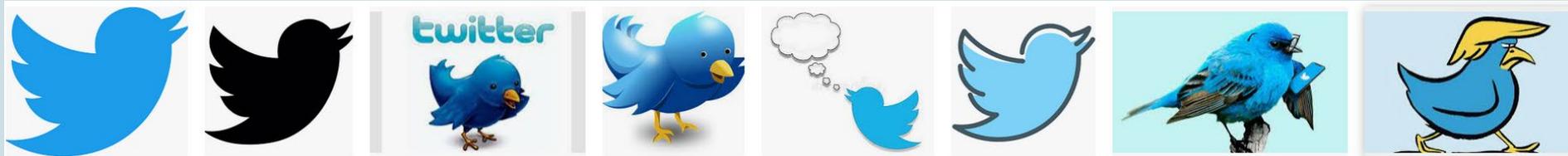
Dürfen wird das alles?

- § 2, Abs. 1 ArchivG NRW: Tweets sind Unterlagen, wenn Social Media zu amtlichen Zwecken eingesetzt wird.
Tweets are documents from a legal point of view
- Urhebergesetz: Schutzfähigkeit für eine Twitter-Meldung („Tweet“) wird verneint, da die notwendige Individualität und Gestaltungshöhe fehlt.
Tweets are no works protected by copyright
- §§ 22-23 Kunsturhebergesetz („Recht am eigenen Bild“): in der Regel unkritisch.
The right to one's own image is unproblematic



An odd duck! – significant properties

Komische Vögel! - Signifikante Eigenschaften



1. Sinnliche Gesamtwahrnehmung / Anmutung
2. „Look-and-Feel“
3. Lesbarkeit
4. Recherchierbarkeit
5. Weiterverwendbarkeit von textlichen Informationen
6. Weiterverwendbarkeit von bildlichen Informationen
7. Weiterverwendbare Video-Inhalte
8. Maschinelle Verarbeitbarkeit der Informationen
9. Versteckte Informationen / detaillierte Metadaten zum Account

1. Overall sensual perception / appearance
2. "Look and Feel"
3. Readability
4. Searchability
5. Reusability of textual information
6. Reusability of pictorial information
7. Reusable video content
8. Machine processability of the information
9. Hidden information / detailed metadata about the account

2020

Projekt in Zusammenarbeit mit der LWL.IT
Project in cooperation with LWL.IT

Günstige Rahmenbedingungen beim LWL

Favourable framework conditions at LWL

- Das Stadtarchiv Münster ist Kunde der LZA-Lösung DiPS.kommunal
- Projekt Web-Archivierung im DiPS-Team der LWL.IT, dabei Evaluation verschiedener Anwendungen und Bibliotheken, u. a. auch TWINT
- Kapazität, ein Werkzeug zu entwickeln

Folge: Entschluss zur gemeinsamen Entwicklung einer Anwendung auf Basis der Bibliothek TWINT

Result: Decision to jointly develop an application based on the TWINT library

- Methodik: Entwicklung eines Prototyps und iterative Optimierung in enger Abstimmung

Twint und seine Eigenschaften

Twint and its features

twint = "Twitter Intelligence Tool"

<https://github.com/twintproject/twint>

- Python-Bibliothek
- Keine Nutzung der Twitter-API
- Triggern der Suchoperatoren von Twitter
- Keine Browser-Emulation
- Basiert auf dem Parser BeautifulSoup
- Ausgabe: textliche Rohdaten eines Suchvorgangs

First steps with Twint

Erste Versuche mit Twint

Markus Lewe
@LeweMarkus

Vier neue E-Busse sind der nächste Schritt der Stadtwerke, Münster ein Stück grüner zu machen. Sie fahren ab diesem Sommer auf der Linie 11 von Gievenbeck durch die Innenstadt bis zur Mondstraße. Die Busse werden mit Ökostrom betankt und fahren leise./tl



4:15 nachm. · 30. Juni 2020 · Twitter Web App

26 „Gefällt mir“-Angaben

Twitterte deine Antwort Antworten

Aimanó @atassi_95 · 30. Juni 2020
Antwort an @LeweMarkus
👍👍

Liveyourdream - Ola ke ola @SabineSeher · 5. Juli 2020
Antwort an @LeweMarkus
Hoppla, da brauchen wir ja bald ein eigenes Kraftwerk...

1277968987150618624	Markus Lewe	2020-06-30	16:15:24	Vier neue E-Busse sind der nächste Schritt der Stadtwerke, Münster ein Stück grüner zu machen. Sie fahren ab diesem Sommer auf der Linie 11 von Gievenbeck durch die Innenstadt bis zur Mondstraße. Die Busse werden mit Ökostrom betankt und fahren leise./tl pic.twitter.com/e5WlQeeHTU	https://pbs.twimg.com/media/EbxA2N8UjAGkW.jpg	
1278082291431954434	Aimanó	2020-06-30	23:47:37	👍👍	lewe/markus	
1279781856342678528	Sabine Seher	2020-07-05	16:19:08	Hoppla, da brauchen wir ja bald ein eigenes Kraftwerk...	lewe/markus	
1277529325802328066	Markus Lewe	2020-06-29	11:08:21	pic.twitter.com/6Ldo0Vgg1G	https://pbs.twimg.com/media/EbyqCDwKAcakWS.jpg	
1278377043240116224	Jutta Manzfeld	2020-07-01	19:16:52	Lieber Herr Lewe, haben Sie ebenfalls eine gute Zeit und bleiben Sie gesund.	lewe/markus	
1276146476612288512	Markus Lewe	2020-06-25	15:33:24	#Spätestich für 247 neue Wohnungen im Zentrum Nord, von denen fast 80% öffentlich gefördert und somit preisgebunden sind. Neben den großen Vorhaben auf den #Konversionsflächen , sind es auch Projekte wie dieses, die zur #Verbesserung der Lage auf dem #Wohnungsmarkt beitragen. /tl pic.twitter.com/HgzCzjQQf	https://pbs.twimg.com/media/EbXlcuXgAAn3NR.jpg https://pbs.twimg.com/media/EbXlcw3K1YADwO6.jpg https://pbs.twimg.com/media/EbXlcw3K1YADwO6.jpg	#spätestich #konversionsflächen #verbesserung #wohnungsmarkt
1276146878845988864	pflegekraft-muenster.de	2020-06-25	15:35:00	Bitte auch an altengerechten Wohnraum denken...	lewe/markus	
1276385736095404034	Markus Lewe	2020-06-26	07:24:08	Machen wir. Ein Teil der Wohnungen wird auch altengerecht ausgestattet sein. Auch Begegnungsräume mit Bereuung sind Teil dieses Projektes. /tl	pflegekraft_ms	

Optimisation: integration of additional libraries

Optimierung und Funktionserweiterung

- Rekursive Erfassung der Antworten durch mehrere parallele Twint-Crawls
- Zuordnung der Antworten anhand der Nutzer und der Chronologie
- Visualisierung der Dialoge über eine eingerückte Darstellung
- Erfassung der Bilder, Icons und Logos mittels der Integration weiterer Bibliotheken
- Einbettung der Bilder in das HTML, um der Ingest-Logik von DiPS.kommunal gerecht zu werden
- Beschleunigung durch Zwischenspeicherung
- Komfortablere Steuerbarkeit über Parameter für Zeitintervalle

Optimisation: integration of additional libraries

Optimierung und Funktionserweiterung

Markus Lewe @LeweMarkus

Vier neue E-Busse sind der nächste Schritt der Stadtwerke, Münster ein Stück grüner zu machen. Sie fahren ab diesem Sommer auf der Linie 11 von Gievenbeck durch die Innenstadt bis zur Mondstraße. Die Busse werden mit Ökostrom betankt und fahren leise./tl



4:15 nachm. · 30. Juni 2020 · Twitter Web App

26 „Gefällt mir“-Angaben

Twitterte deine Antwort Antworten

Aimanó @atassi_95 · 30. Juni 2020
Antwort an @LeweMarkus
👍👌

Liveyourdream - Ola ke ola @SabineSeher · 5. Juli 2020
Antwort an @LeweMarkus
Hoppla, da brauchen wir ja bald ein eigenes Kraftwerk...

Juni 2020 6 Tweets

Markus Lewe lewemarkus 1277968987150618624
30.06.2020, 16:15:24

Vier neue E-Busse sind der nächste Schritt der Stadtwerke, Münster ein Stück grüner zu machen. Sie fahren ab diesem Sommer auf der Linie 11 von Gievenbeck durch die Innenstadt bis zur Mondstraße. Die Busse werden mit Ökostrom betankt und fahren leise./tl <https://t.co/e5WlQeeHTU>



1277968987150618624.EbxA2N8U8AYJGkKW.jpg

29 „Gefällt mir“-Angaben

Aimanó atassi_95 1278082791431954434
30.06.2020, 23:47:37 · Antwort an LeweMarkus
@LeweMarkus 👍👌

Sabamed Ola ke ola sabineseher 1279781866342678528
05.07.2020, 16:19:08 · Antwort an LeweMarkus
@LeweMarkus Hoppla, da brauchen wir ja bald ein eigenes Kraftwerk...

Generated Submission Information Packages

Die erzeugten SIP

Python-
Skript
„twitter-
report.py“

...
by the Twint
Python script

Ergänzend

+
Additional
information

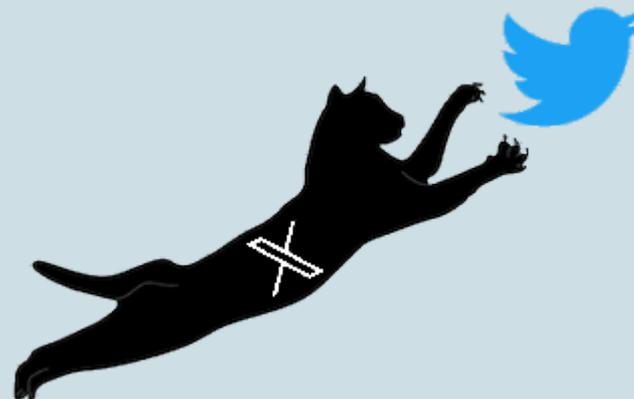
```
data.json
twitter-report.lewemarkus.2020-10-30.html
982665602602012672.DaMhHvTW4AEfXnc.jpg
982670227057061888.DaMlSpUX4AAN2Hr.jpg
982670227057061888.DaMlWSFW4AE0KLy.jpg
984148128919105536.Dahlel1XUAE552-.jpg
...
hashtag-list.lewemarkus.2020-10-30.txt
```

```
Screenshot_Beispieltweet_Lewe.JPG
Screenshot_Profil_Lewe.JPG
Screenshot_Tweets_Lewe.JPG
Follower_Lewe_User.xlsx
Following_Lewe_User.xlsx
```

Oktober 2022

Übernahme durch Elon Musk und Paradigmenwechsel bei Twitter

Takeover by Elon Musk and paradigm shift at Twitter



Schrittweiser Umbau der Plattform

Successive conversion of the platform

Ab Februar 2023

- Weitgehende Modifikationen in der Technik der Website
- Einführung kostenpflichtiger Funktionen

Juni 2023

- Einführung eines Limits für die abzurufenden Tweets
- Integration von Mechanismen zum Aussperren von Crawlern
- Zwang zum Login

Juli 2023

- Umbenennung zu X

Folgen des Umbaus für unser Werkzeug

Consequences for our tool

- Anfangs noch Reaktionen in der Twint-Community, die neuen Hürden zu überwinden
- Kapitulation vor den technischen Hürden: Einstellung von Twint
- Somit keine Möglichkeit mehr, Daten per Crawl abzurufen.

Folge: Das Werkzeug `twitter-report.py` ist im März 2023 funktionslos geworden

Ausblick

**Neubewertung und Suche nach neuen Lösungsansätzen
zur Social-Media-Archivierung**

**Prospects, new evaluation and search for new solutions
for social media archiving**

Wandel in der Social-Media-Welt

Change in the world of social media

- Unseriöse Inhalte und kostenpflichtige Modelle führen bei X zu einer Abkehr der Nutzer, das Image der Marke ist beschädigt
- Sinkende Nutzerzahlen bei X bewirken sinkende Werbeerlöse, woraufhin weitere Ansätze zur Monetarisierung verfolgt werden
- Multiple neue Anbieter und Plattformen entstehen, aber keine vollwertige Alternative bildet sich heraus: Mastodon-Instanzen, Bluesky, Threads, Pebble...

Situation from an archiving point of view

Situation aus Perspektive der Archivierung



„Der OB entscheidet, dass die Stadt nicht mehr im Social-Media-Kanal X (ehemals Twitter) vertreten sein soll und [dieser] nicht weiter betrieben wird.“

Beschluss im
Verwaltungsvorstand am
19.12.2023

Situation aus Perspektive der Technik

Situation from a technical point of view

Erweiterung des Werkzeugkastens zur Web-Archivierung: Evaluation neuer Werkzeuge für Social-Media

Geeignet erscheinen Werkzeuge mit integrierter Browser-Engine.
Evaluiert werden:



Brozzler



webrecorder/browsertrix



Archivfachliche Überlegungen und erste Testläufe:

Peter Worm, Twitter Accounts von kommunalen Wahlbeamten.

Bewertung und Überlieferungsstrategien. In: Archive 2.0 am 03.02.2020

<https://archive20.hypotheses.org/8710>

Die Twint-Lösung und ein How-To:
Michael Jerusalem, Marc Malwitz, Peter Worm: Ein neuer Ansatz für die Langzeitarchivierung von Twitter-Accounts. In: Archive 2.0 am 09.03.2021

<https://archive20.hypotheses.org/10031>

Vielen Dank!

Thank you for your attention!