



Herausforderung Social- Media-Archivierung. Archivierungsstrategie von Facebook im Stadtarchiv Biberach. Ein Werkstattbericht.

Referentin: Stefanie
Hartmannsgruber

Agenda

- f Die archivische Bewertung**
- f Die Archivierungsstrategie**
- f Ein Ausblick**

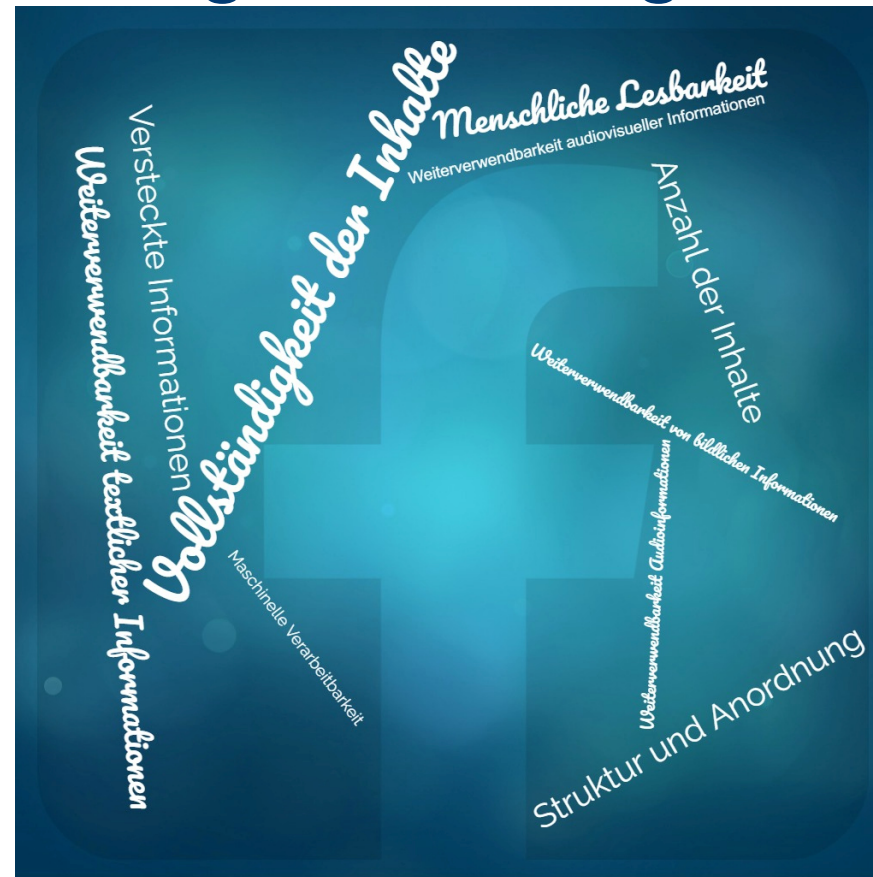


Social-Media als Archivquelle?

Das vierphasige Bewertungsmodell



Exkurs: signifikante Eigenschaften



Dürfen wir die Inhalte aus Facebook archivieren?

Die rechtlichen Problemfelder

Datenbesitz

Urheberrecht

Persönlichkeits-
recht

Dürfen wir die Inhalte aus Facebook archivieren?

Der Lösungsansatz

Datenbesitz

§ 312f Abs. 3 BGB

Urheberrecht

§ 60f Abs 2 UrhG in
Kombination mit § 7 Abs. 2
LarchG BW sowie
§ 2 Abs. 2 LarchG BW

Persönlichkeits-
recht

§ 6 Abs. 1e DSGVO in
Kombination mit § 14
Abs. 1 LDSG BW sowie §
7 Abs. 2 LarchG BW

Wie können wir Inhalte aus Facebook archivieren?

JSON als Archivierungsformat

```
1 [
2   {
3     "post_id": "847600947405219",
4     "text": "🚧 Die Scheffelstraße auf dem Mittelberg wird komplett erneuert. Die Bauarbeiten beginnen heute.👷 Die Scheffelstraße gehört zu den schlechtesten Straßen Biberachs. Es muss nicht nur die Fahrbahnfläche, sondern auch der Stra\u00dfenuntergru\u00df verbessert werden. Daher reicht eine Sanierung nicht aus. Sie..",
5     "post_text": "🚧 Die Scheffelstraße auf dem Mittelberg wird komplett erneuert. Die Bauarbeiten beginnen heute.👷 Die Scheffelstraße gehört zu den schlechtesten Straßen Biberachs. Es muss nicht nur die Fahrbahnfläche, sondern auch der Stra\u00dfenuntergru\u00df verbessert werden. Daher reicht eine Sanierung nicht aus. Sie..",
6     "shared_text": "",
7     "original_text": null,
8     "time": "2024-03-13 15:00:00",
9     "timestamp": null,
10    "image": "https://scontent.fstr1-1.fna.fbcdn.net/v/t39.30808-6/432433711_847600910738556_1619170538783076121_n.jpg?stp=cp0_dst-jpg_e15_p320x320_q65&nc_cat=104&ccb=1-7&nc_sid=5f2048&efg=eyJpIjojYiJ9&nc_ohc=sBNsBmXLGzMAX-oxw72&nc_oc=AQlwEfq61JC2zTxaQETL0rCXnA33kxBNt5fP-MGLHsI4L7T9K8gU35yIrjMCbraWSac&nc_ht=scontent.fstr1-1.fna&oh=00_AfDyUhJmUdw-kDKtG0m-P4yFtUHDyrY87ch84sYKGUd8Lw6oe=65F82166",
11    "image_lowquality": "https://scontent.fstr1-1.fna.fbcdn.net/v/t39.30808-6/432433711_847600910738556_1619170538783076121_n.jpg?stp=cp0_dst-jpg_e15_p206x206_q65&nc_cat=104&ccb=1-7&nc_sid=5f2048&efg=eyJpIjojYiJ9&nc_ohc=sBNsBmXLGzMAX-oxw72&nc_oc=AQlwEfq61JC2zTxaQETL0rCXnA33kxBNt5fP-MGLHsI4L7T9K8gU35yIrjMCbraWSac&nc_ht=scontent.fstr1-1.fna&oh=00_AfAtudIhTf2shLEB2p7YPqojjCaZ6qaQDgDigPedaoqrhQ&oe=65F82166"
```

Wie können wir Inhalte aus Facebook archivieren?

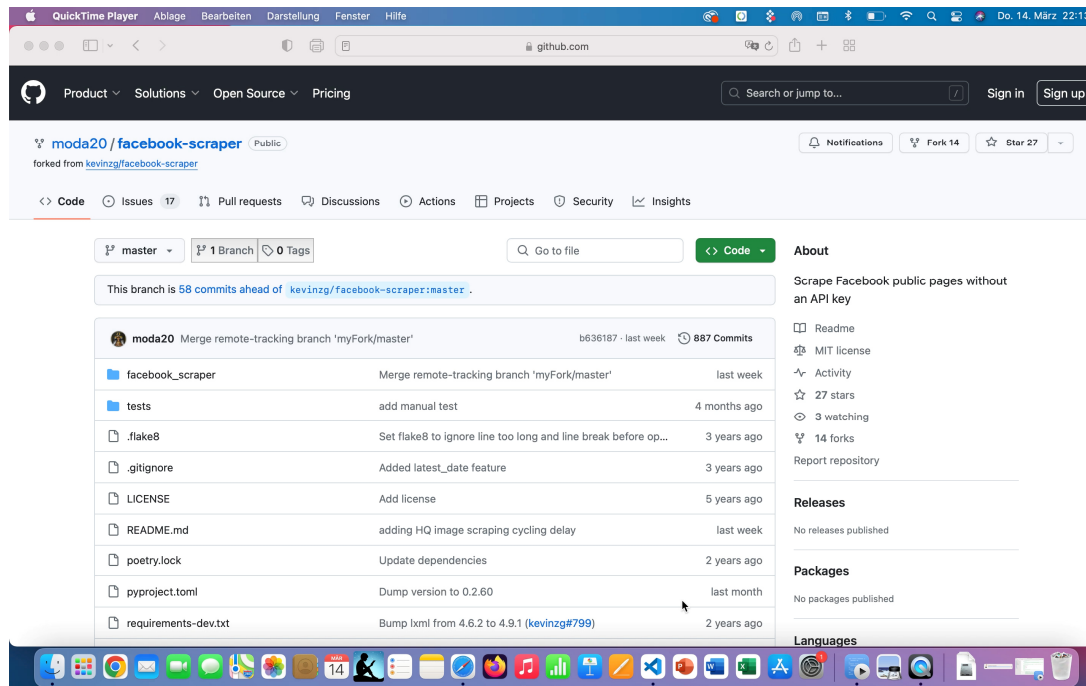
Gegenüberstellung von Methoden

Social-Graph-API
mit Facebook-Scraper

Dienstleister

Facebook Exportfunktion

Unser praktischer Lösungsansatz Der Social-Graph-API mit Facebook-Scraper

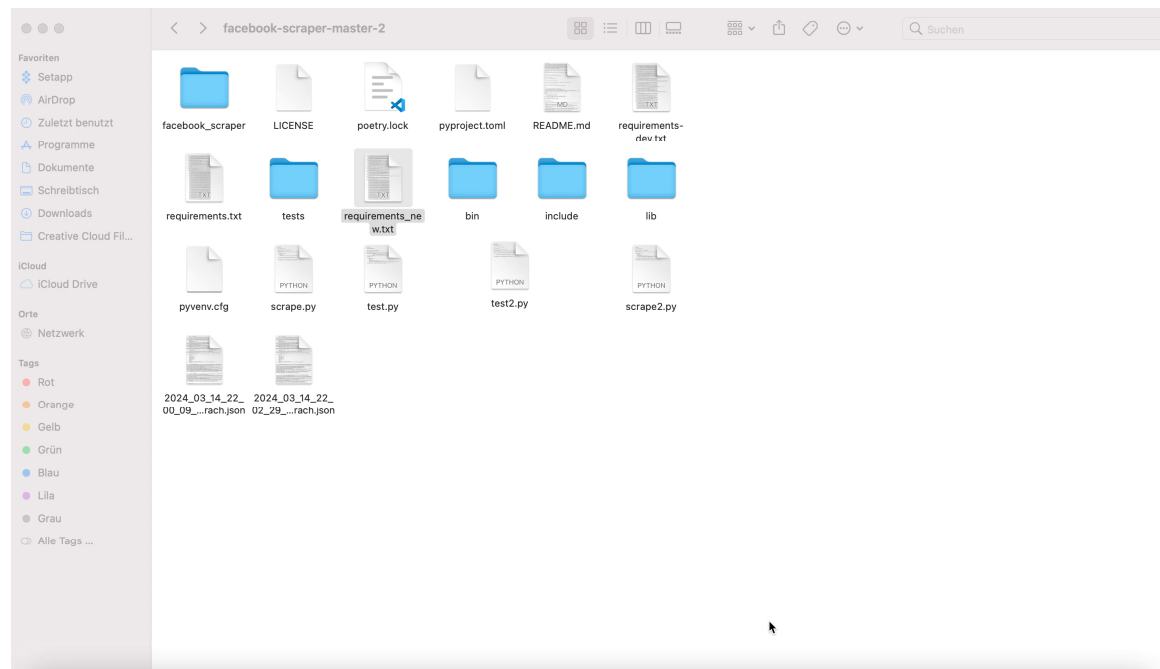


The screenshot shows a GitHub repository page for 'moda20/facebook-scraper'. The repository is public and has 14 forks and 27 stars. The main branch is 'master', which is 58 commits ahead of the original repository. The repository contains several files and folders, including 'facebook_scraper', 'tests', '.flake8', '.gitignore', 'LICENSE', 'README.md', 'poetry.lock', 'pyproject.toml', and 'requirements-dev.txt'. The 'pyproject.toml' file was last updated last month, and 'requirements-dev.txt' was updated 2 years ago. The repository is described as a scraper for Facebook public pages without an API key. The page also includes sections for 'About', 'Releases', and 'Packages', all of which are currently empty.

File/Folder	Description	Last Updated
facebook_scraper	Merge remote-tracking branch 'myFork/master'	last week
tests	add manual test	4 months ago
.flake8	Set flake8 to ignore line too long and line break before op...	3 years ago
.gitignore	Added latest_date feature	3 years ago
LICENSE	Add license	5 years ago
README.md	adding HQ image scraping cycling delay	last week
poetry.lock	Update dependencies	2 years ago
pyproject.toml	Dump version to 0.2.60	last month
requirements-dev.txt	Bump lxml from 4.6.2 to 4.9.1 (kevinzg#799)	2 years ago

Der Facebook-Scraper

Voraussetzung und Umgebung



Der Facebook-Scraper

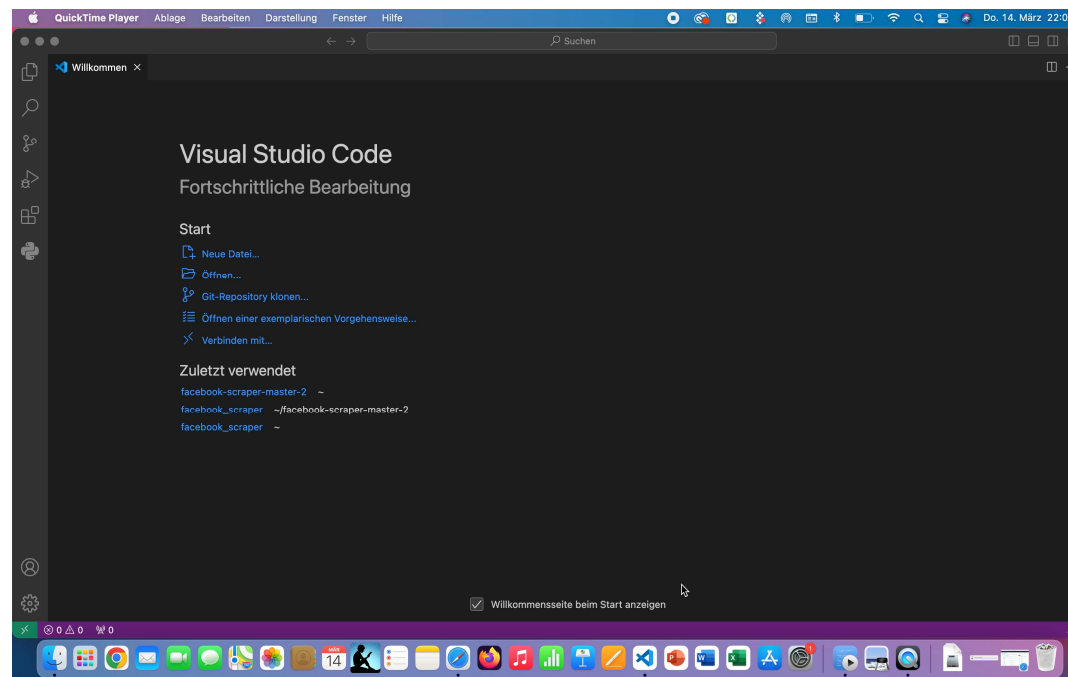
Das Python-Script

```
15 ## Usage
16
17 in order to get everything running right, follow these steps
18
19 1. Send the unique **page name, profile name, or ID** as the first parameter
20 2. Specify the base_url and start_url to use the mbasic page instead
21 3. Get the mbasicHeaders that you want to use and read them from a file in order to inject them into the scraper.
22 **you can get these headers from opening an example page and selecting a high-end device in the developer tools (such as samsung s20
23 ultra)
24 . This will help with getting newer versions of posts and higher fidelity images.**
```

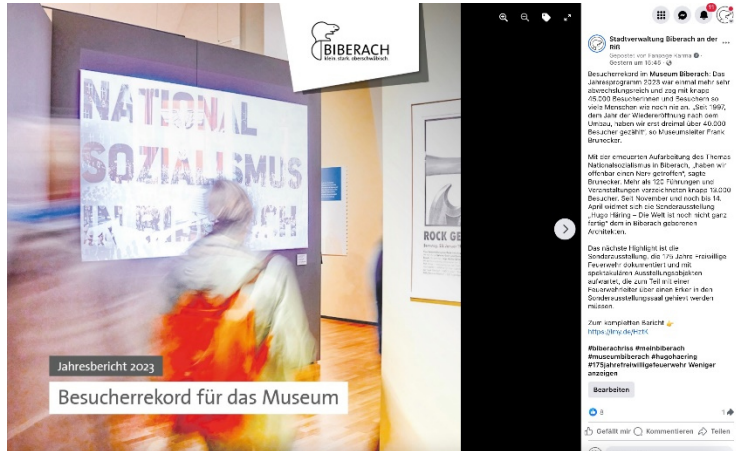
```
1 import json
2 import logging
3 from datetime import datetime
4
5 import facebook_scraper
6
7 # facebook_scraper.enable_logging(logging.DEBUG)
8
9 name_of_page = "biberach"
10 options={"comments": True, "reactors": True, "progress": True, "allow_extra_requests": True}
11
12 posts_list = list(
13     facebook_scraper.get_posts(
14         name_of_page,
15         pages=1,
16         options=options,
17         base_url="https://mbasic.facebook.com",
18         start_url="https://mbasic.facebook.com/biberach?v=timeline",
19         cookies="/Users/stefaniehartmannsgruber/Documents/Facebook-Scraper Tobi/Cookies/www.facebook.com_cookies.txt"
20 )
21 # for post in get_posts(name_of_page, pages=2, options=options):
22 #     print(f"Comments: {post.get('comments')}, Likes: {post.get('likes')}")
23
24 with open(datetime.now().strftime("%Y_%m_%d_%H_%M_%S") + "_" + name_of_page + ".json", "w") as f:
25     json.dump(posts_list, f, indent=4, default=str, ensure_ascii=False)
```

Der Facebook-Scraper

Die Ausführung



Der Facebook-Scraper Die Dateiausgabe



```

160 {
161   "post_id": "847859324126046",
162   "text": "Besucherrekord im Museum Biberach: Das Jahresprogramm 2023 war einmal mehr sehr abwechslungsreich und zog mit knapp 45.000 Besucherinnen und Besuchern so viele Menschen wie noch nie an. „Seit 1997, dem Jahr der Wiedereröffnung nach dem Umbau, haben wir erst dreimal über 40.000 Besucher gezählt“, so...“
163   "post_text": "Besucherrekord im Museum Biberach: Das Jahresprogramm 2023 war einmal mehr sehr abwechslungsreich und zog mit knapp 45.000 Besucherinnen und Besuchern so viele Menschen wie noch nie an. „Seit 1997, dem Jahr der Wiedereröffnung nach dem Umbau, haben wir erst dreimal über 40.000 Besucher gezählt“, so...“
164   "shared_text": "",
165   "original_text": null,
166   "time": "2024-03-13 15:45:00",
167
168   },
169   "video": null,
170   "video_duration_seconds": null,
171   "video_height": null,
172   "video_id": null,
173   "video_quality": null,
174   "video_size_MB": null,
175   "video_thumbnail": null,
176   "video_watches": null,
177   "video_width": null,
178   "likes": 8,
179   "comments": 8,
180   "shares": 0,
181   "post_url": "https://facebook.com/story.php?story_fbid=pfbid8eXRZvkrR44kLJwQ3HLKt9ARx81Ur3Pi3VT51nwK4pKmYQ7DYMg4WqDXZTSYsr1cd16id=100864659640712",
182   "link": null,

```

Die Archivierungsdatei

Das SIP

Der Content
(Posts, Reactions, Comments)

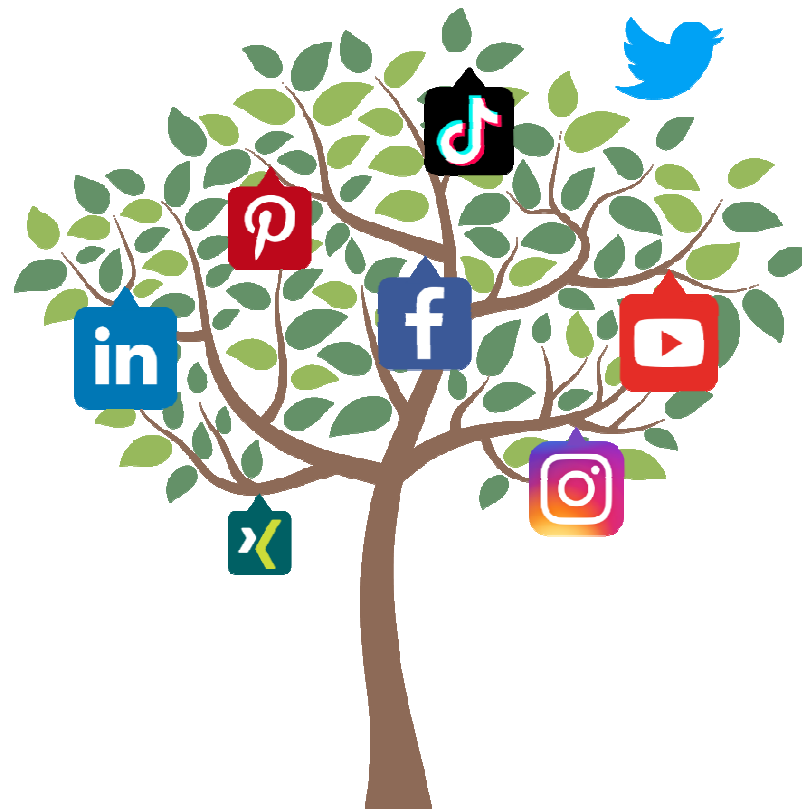
Die Reichweite
(Insight Auswertung)

Die Profilinformationen
(versteckte Informationen)

Ausblick

Weiter-
entwicklun
g Skript
und
Programm

Andere
Social-
Media
Dienste



Benutzung
und
Erschließung
der Daten

Facebook
Auftritte
Dritter

Kontakt

Stefanie Hartmannsgruber

Sachbearbeiterin für Digitale Langzeitarchivierung und Digitalisierung

Stadtarchiv Biberach

Waldseer Straße 31

88400 Biberach

Tel. 07351/51-625

E-Mail: s.hartmannsgruber@biberach-riss.de



Bildnachweis: Titelfolie

Abbildung: Social-Media-Netzwerk (Designer: The Digital Artist)

<https://pixabay.com/de/illustrations/facebook-social-media-netzwerk-3391214/>

Bildnachweis: Foliennummer 2

Video: Quellcode für das WWW <https://www.sothebys.com/en/buy/auction/2021/this-changed-everything-source-code-for-www-x-tim-berners-lee-an-nft/source-code-for-the-www>

Bildnachweis: Foliennummer 4

Abbildung: Stadt- und Stiftungsarchiv Aschaffenburg (Fotograf: Till Benzin) [https://stadtarchiv-aschaffenburg.de/das-archiv#iLightbox\[c9afc76211c092ea999\]/0](https://stadtarchiv-aschaffenburg.de/das-archiv#iLightbox[c9afc76211c092ea999]/0)

Abbildung: Library of Congress Washington (Fotograf: Kqcycler) <https://pixabay.com/de/photos/us-library-of-congress-bibliothek-3762072/>

Abbildung: The National Archives UK (Fotograf: Jim Linwood) <https://artuk.org/visit/venues/the-national-archives-7568>

Abbildung: Stadtarchiv Münster (Fotograf: unbekannt) <https://www.archive.nrw.de/stadtarchiv-muenster>

Abbildung: Twitter Logo (Grafiker: Clker Free Vector Images) <https://pixabay.com/de/vectors/twitter-twitter-vogel-312464/>