



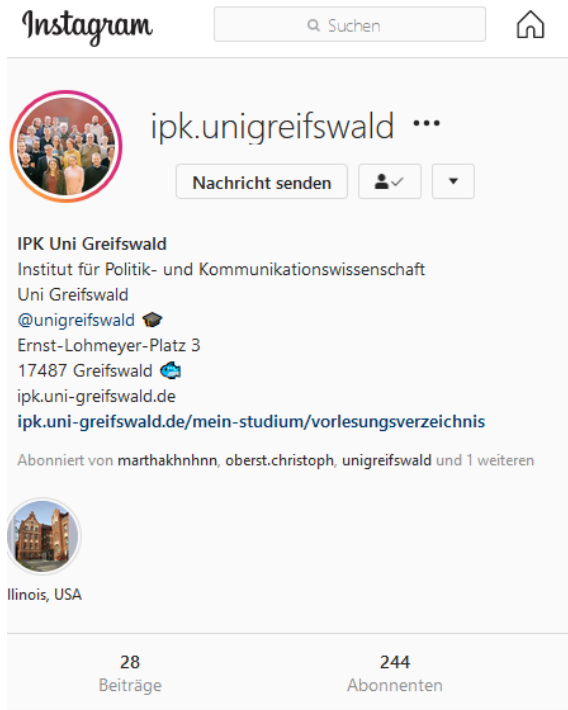
Automatisierte Erhebung von Social Media Daten

Workshop Teil II: APIs – Datenerhebung mit
Facepager

Datenzugänge:

https://www.instagram.com/ipk.unigreifswald/?__a=1

User Interface (Browser)



HTML (Web scraping)

```
1 <!DOCTYPE html>
2 <html lang="de" class="no-js logged-in clier
3   <head>
4     <meta charset="utf-8">
5     <meta http-equiv="X-UA-Compatible" c
6
7     <title>
8 IPK Uni Greifswald (@ipk.unigreifswald) • Ir
9 </title>
10
11
12 <meta name="robots" content="noimage
13 <meta name="apple-mobile-web-app-sta
14 <meta name="mobile-web-app-capable"
15 <meta name="theme-color" content="#f
16 <meta id="viewport" name="viewport"
17 <link rel="manifest" href="/data/mar
18
19 <link rel="preload" href="/static/bu
20 <link rel="preload" href="/static/bundles/es
21 <link rel="preload" href="/static/bundles/es
22 <link rel="preload" href="/static/bundles/es
23 <link rel="preload" href="/static/bundles/es
24 <link rel="preload" href="/static/bundles/es
25 <link rel="preload" href="/static/bundles/es
26 <link rel="preload" href="/static/bundles/es
27 <link rel="preload" href="/static/bundles/es
28 <link rel="preload" href="/static/bundles/es
29 <link rel="preload" href="/static/bundles/es
30
31
32 <script type="text/javascript">
33 (function() {
34
35 var docElement = document.documentElement;
36
```

JSON (API)

```
{
  logging_page_id: "profilePage_23191765473",
  show_suggested_profiles: false,
  show_follow_dialog: false,
  graphql: {
    user: {
      biography: "Institut für Politik- und  
Kommunikationswissenschaft\nUni Greifswald\n@un  
Ernst-Lohmeyer-Platz 3\n17487 Greifswald  
greifswald.de",
      blocked_by_viewer: false,
      restricted_by_viewer: false,
      country_block: false,
      external_url: https://ipk.uni-greifswald.de/mei  
/vorlesungsverzeichnis/,
      external_url_linkshimmed: https://l.instagram.c  
%2Fipk.uni-greifswald.de%2Fmein-  
studium%2Fvorlesungsverzeichnis%2Fse=ATMEwB9fx5  
\_0ypUxMLK3fmyoA7pz\_vMNTCmkjBAX-  
CXX7Y0Ap0tyqvzDTxAURk404W81lX9X79sdNRW&s=1,
      edge_followed_by: {
        count: 244
      },
      followed_by_viewer: true,
      edge_follow: {
        count: 22
      },
      follows_viewer: false,
      full name: "IPK Uni Greifswald",
      has_ar_effects: false,
      has_channel: false,
      has_blocked_viewer: false,
      highlight_reel count: 1,
      has_requested_viewer: false,
      id: "23191765473",
      is business account: true,
    }
  }
}
```

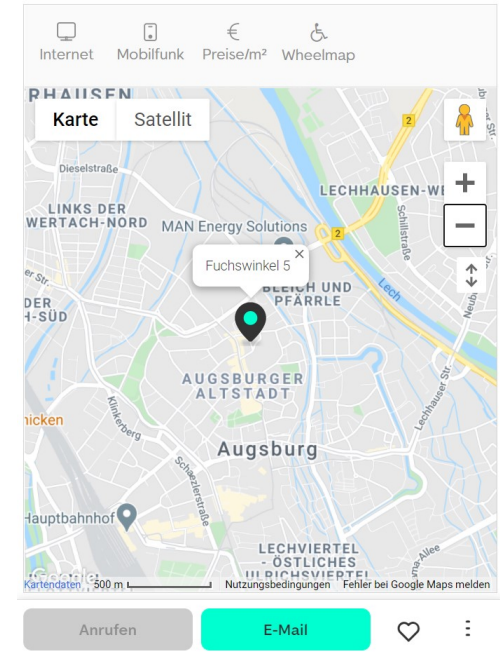


APIs

- Werden von Betreibern von Websites bereitgestellt
- Programmierschnittstellen, über die zwei Programme miteinander interagieren können (Jacobson, Brail & Woods 2012: 5).
- Für Wissenschaft: Vorstrukturierter Datenzugang (oftmals JSON)
- Häufig mit Zugriffsbeschränkung (Authentifizierung, Rate-Limits)

Karte

Fuchswinkel 5, 86152 Augsburg, Bleich und Pfärrle
[Mehr Informationen zu dieser Adresse](#)



Quelle: Immobilienscout24.de



Quelle: Zeit.de

Dienste und Plattformen mit APIs (Auswahl)

Social Media

- Facebook (<https://developers.facebook.com/>)
- Twitter (<https://developer.twitter.com/en/docs>)
- YouTube (<https://developers.google.com/youtube/v3>)
- Flickr (<https://www.flickr.com/services/api/>)
- Reddit (<https://www.reddit.com/dev/api/>)
- LinkedIn (<https://www.linkedin.com/developers/>)

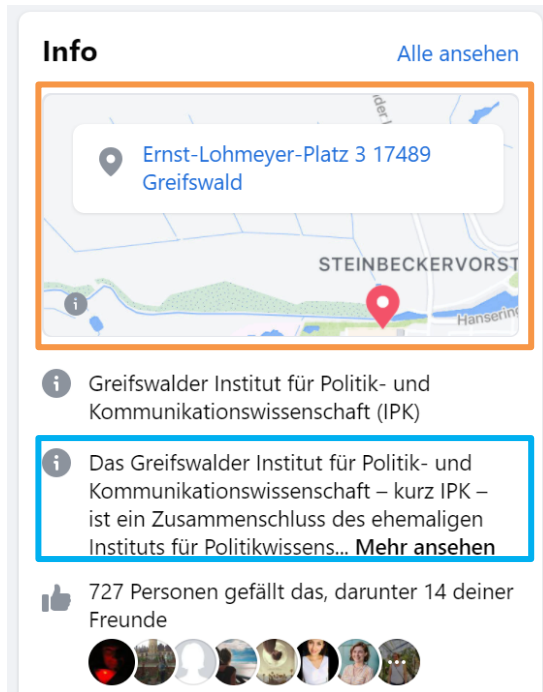
Messenger: Telegram, WhatsApp, Threema, Skype, Discord

Streaming: Podcast.de, Spotify, Apple Music, Vimeo, Twitch

Andere Dienste: Google Maps, Amazon,

Facebook-Seite des IPK Greifswald im Browser und als JSON

Browser-Anzeige



JSON

```
{
  "id": "288837164503756",
  "name": "Greifswalder Institut für Politik- und Kommunikationswissenschaft (IPK)",
  "location": {
    "city": "Greifswald",
    "country": "Germany",
    "latitude": 54.098440810237,
    "located_in": "234989956684237",
    "longitude": 13.375945687294,
    "street": "Ernst-Lohmeyer-Platz 3",
    "zip": "17489"
  },
  "about": "Greifswalder Institut für Politik- und Kommunikationswissenschaft (IPK)",
  "description": "Das Greifswalder Institut für Politik- und Kommunikationswissenschaft - kurz IPK - ist ein Zusammenschluss des ehemaligen Instituts für Politikwissenschaft und des zuvor in der Deutschen Philologie angesiedelten Lehrstuhls für Kommunikationswissenschaft. Das so entstandene Institut ist das bislang einzige seiner Art in Deutschland.",
  "fan_count": 727
}
```

Likes der IPK-Facebook-Seite als JSON und Tabelle

JSON

```
"data": [  
  {  
    "id": "145718628797076",  
    "name": "webmoritz.de",  
    "fan_count": 1470  
  },  
  {  
    "id": "340018314585",  
    "name": "Greifswald international",  
    "fan_count": 2732  
  },  
  {  
    "id": "182566151762895",  
    "name": "Fennistik UniGreifswald",  
    "fan_count": 1051  
  },  
  {  
    "id": "139610749421232",  
    "name": "BCW Deutschland",  
    "fan_count": 665  
  }  
],
```

Tabelle

id	name	fan_count
145718628797076	webmoritz.de	1470
340018314585	Greifswald international	2732
182566151762895	Fennistik UniGreifswald	1051
139610749421232	BCW Deutschland	665

Abfragen der Facebook-API: Zusammensetzung von URLs

Pfad zur Ressource, die abgefragt wird
z. B. Feed einer Seite,
Kommentare eines Posts,
Likes eines Kommentars.

`https://graph.facebook.com/v3.2/Endpunkt&Parameter`

Basispfad der API

Zusätzliche Einstellungen
z.B. Sortierung nach Datum

Abfragen von Likes von Facebook-Seiten

Endpunkt=
Pfad zur Ressource

https://graph.facebook.com/v3.2/IPK.Greifswald/likes

Basispfad

Automatisierung → viele URLs gleichzeitig zusammensetzen und abfragen



- **Automatisierte Datenerhebung** mittels APIs und Webscraping ohne Programmierung
- **Entwicklung** seit 2012 von Till Keyling (München) und Jakob Jünger (Greifswald)
- **Open Source Project** für Windows und Mac:
<https://github.com/strohne/Facepager>



- **API-Module:** YouTube, Twitter, Facebook, Amazon, Generic. Abfrageparameter frei einstellbar
- **Presets** zur Dokumentation, zum Austausch von Erhebungsdesigns und zum schnellen Einstieg
- **Exportieren** als CSV-Datei oder über die Zwischenablage
- **Hilfe** über die eingebaute API-Dokumentation, Wiki und Facebook-Gruppe

Layout von Facepager

The screenshot shows the Facepager 4.3 application interface. The components are labeled as follows:

- Menu Bar:** Located at the top right, containing window control buttons.
- Data View:** Located in the top right pane, showing a table with columns 'Key' and 'id', and a single row with values 'name' and '340018314585'.
- Nodes View:** Located in the top left pane, showing a table with columns 'Object ID', 'Object Type', 'Query Status', 'Query Time', and 'Query Type'. It lists several data points for 'IPK.Greifswald'.
- Column Setup:** Located in the middle right pane, showing a table with columns 'id', 'name', and 'fan_co'.
- Query Setup:** Located in the bottom left pane, showing fields for 'Base path', 'Resource', 'Parameters', 'Maximum pages', and 'Access token', along with 'Settings' and 'Login to Facebook' buttons.
- Status View:** Located in the bottom right pane, showing a 'Status Log' with a list of timestamps and actions.

The bottom status bar shows the file path 'C:/Users/Chantal/Documents/Workshop/IPK.db' and the message 'Timer stopped 1 node(s) selected'.

Input

Facepager 4.3

Open Database New Database Export Data Add Nodes Delete Nodes Presets APIs Help

Expand nodes Collapse nodes Find nodes Copy Nodes to Clipboard

Add Column Add All Columns Extract Data

Object ID	Object Type	Query Status	Query Time	Query Type
IPK.Greifswald	seed			
1		fetchd (200)		page-id>/like
3		fetchd (200)		page-id>/like
1		fetchd (200)		page-id>/like
1		fetchd (200)		page-id>/like
510585058950799	data	fetchd (200)		page-id>/like
116277705061155	data	fetchd (200)	2021-03-05 18:...	Facebook:<page-id>/like
112115572230196	data	fetchd (200)	2021-03-05 18:...	Facebook:<page-id>/like
252163088650	data	fetchd (200)	2021-03-05 18:...	Facebook:<page-id>/like

1. Neue Datenbank

2. Knoten einfügen

Key	Value
name	Greifswald international
id	340018314585

Custom Table Columns (one key per line)

id
name
fan_count

Apply Column Setup Clear Columns

YouTube Twitter Twitter Streaming Facebook Amazon Generic

Base path

Resource

Parameters

Maximum pages

Access token

Settings

Node level

Select all nodes ☐

Exclude object types

Resume collection ☐

Parallel Threads

Requests per minute

Maximum errors

More settings ...

Fetch Data

Status Log

2021-03-05 18:44:53.391039 Start fetch
2021-03-05 18:44:53.417359 Added
2021-03-05 18:44:53.525071 Fetching
2021-03-05 18:44:53.803652 Fetching

C:/Users/Chantal/Documents/Workshop/IPK.db

Timer stopped 1 node(s) selected

Throughput

Facepager 4.3

Open Database New Database Export Data Add Nodes Delete Nodes Presets APIs Help

Expand nodes Collapse nodes Find nodes Copy Node(s) to Clipboard Add Column Add All Columns Extract Data

Object ID	Object Type	Query Status	Query Time	Query Type
IPK.Greifswald	seed			
145718628797076	data	fetchd (200)	2021-03-05 18:44:53.391039	Facebook:/<page-id>/likes
340018314585	data	fetchd (200)	2021-03-05 18:44:53.417359	Facebook:/<page-id>/likes
182566151762895	data	fetchd (200)	2021-03-05 18:44:53.525071	Facebook:/<page-id>/likes
139610749421232	data	fetchd (200)	2021-03-05 18:44:53.803652	Facebook:/<page-id>/likes

3. Preset laden

4. Einstellungen anpassen

5. Daten abfragen

YouTube Twitter Twitter Reaming Facebook Amazon Generic

Base path:

Resource:

Parameters:

Maximum pages:

Access token:

Settings

Node level:

Requests per minute:

Maximum errors:

More settings:

Fetch Data

Status Log

2021-03-05 18:44:53.391039 Start fetch
2021-03-05 18:44:53.417359 Added
2021-03-05 18:44:53.525071 Fetching
2021-03-05 18:44:53.803652 Fetching

C:/Users/Chantal/Documents/Workshop/IPK.db

Timer stopped 1 node(s) selected

z.B.: Einstellungen zum Abfragen der Likes der IPK Facebook-Seite

Expand nodes Collapse nodes Find nodes Copy Node(s) to Clipboard

Object ID	Object Type	Query Status	Query Time	Query Type
IPK.Greifswald	seed			

< >

YouTube Twitter Twitter Streaming Facebook Amazon Generic

Base path

Resource

Parameters

Maximum pages

Access token

Settings Login to Facebook

<https://graph.facebook.com/v3.2/IPK.Greifswald/likes>

Output

Facepager 4.3

Open Database New Database Export Data Add Nodes Delete Nodes Presets APIs Help

Expand nodes Collapse nodes Find node Copy Node(s) to Clipboard Add Column Add All Columns Extract Data

Object ID	Object Type	Query Status	Query Time	Query Type
14571862		(200)	2021-03-05 18:...	Facebook:/<page-id>/like
34001831		(200)	2021-03-05 18:...	Facebook:/<page-id>/like
182566151762895	data	fetchd (200)	2021-03-05 18:...	Facebook:/<page-id>/like
139610749421232	data	fetchd (200)	2021-03-05 18:...	Facebook:/<page-id>/like
510585658956799	data	fetchd (200)	2021-03-05 18:...	Facebook:/<page-id>/like
116277705061155	data	fetchd (200)	2021-03-05 18:...	Facebook:/<page-id>/like
112115572230196	data	fetchd (200)	2021-03-05 18:...	Facebook:/<page-id>/like
252163088650	data	fetchd (200)	2021-03-05 18:...	Facebook:/<page-id>/like

8. Exportieren

6. Daten inspizieren

Key	Value
name	Greifswald international
id	340018314585

Custom Table Columns (one key per line)

id
name
fan_count

Apply Column Setup Clear Columns

YouTube Twitter Twitter Streaming Facebook Amazon Generic

Base path

Resource

Parameters

Maximum pages

Access token

Settings Login to Facebook

Node level

Select all nodes ☐

Exclude object types

Resume collection ☐

Parallel Threads

Requests per minute

Maximum errors

More settings ...

Fetch Data

Status Log

2021-03-05 18:44:18.391039 Start fel

7. Spalten anpassen

C:/Users/Chantal/Documents/Workshop/IPK.db

Timer stopped 1 node(s) selected

Übung in Facepager