



TECHNIK
HOCHSCHULE MAINZ
UNIVERSITY OF
APPLIED SCIENCES

Räumliche Verortung von textbasierten Social-Media-Einträgen am Beispiel von Polizei-Tweets

Svenja Ruthmann
Alexander Rolwes
Klaus Böhm

Lauchingen

Polizei Freiburg @PolizeiFR · 22 Std.
 Lauchingen: Couragierte Frau hindert betrunkene 66-jährige Autofahrerin an Weiterfahrt. #Alkoholüberprüfung der #Polizei ergibt 2,5 Promille. Meldung: bit.ly/2SlumiR



Lauchingen: Couragierte Frau hindert betrunkene 66-jährige Autofahrerin an Weiterfahrt

1 4

Polizei Freiburg @PolizeiFR · 4 Std.
 #Waldshut-#Tiengen: Überfall auf 59-jährigen Mann - Zeugensuche: presseportal.de/blaulicht/pm/1...

Hinweise bitte an die #Kripo Waldshut unter ☎ 07741 8316-0



Waldshut-Tiengen: Überfall auf 59-jährigen Mann

3

Waldshut-Tiengen

Waldshut

Titisee-Neustadt

Freiburg

Polizei Freiburg @PolizeiFR · 21 Std.
 #Titisee-#Neustadt: Männliche Person tot aus Bach geborgen - Bisher keine Hinweise auf Fremdverschulden - #Polizei sucht mögliche Zeugen: presseportal.de/blaulicht/pm/1...

Die Polizei bittet Zeugen sich bei der #Kripo #Freiburg unter ☎ 0761 882 5777 zu melden.



4 1

Polizei Freiburg @PolizeiFR · 1 Std.
 #Gundelfingen: Geisterfahrer auf #B3 zwischen #Emmendingen und #Freiburg - Die #Polizei sucht Zeugen: presseportal.de/blaulicht/pm/1...

Zeugen werden gebeten sich beim Polizeirevier Freiburg Nord ☎ 0761 882-4221 zu melden.



Gundelfingen: Geisterfahrer auf B3 zwischen Emmendingen und Freiburg

1

Gundelfingen

Emmendingen

Freiburg

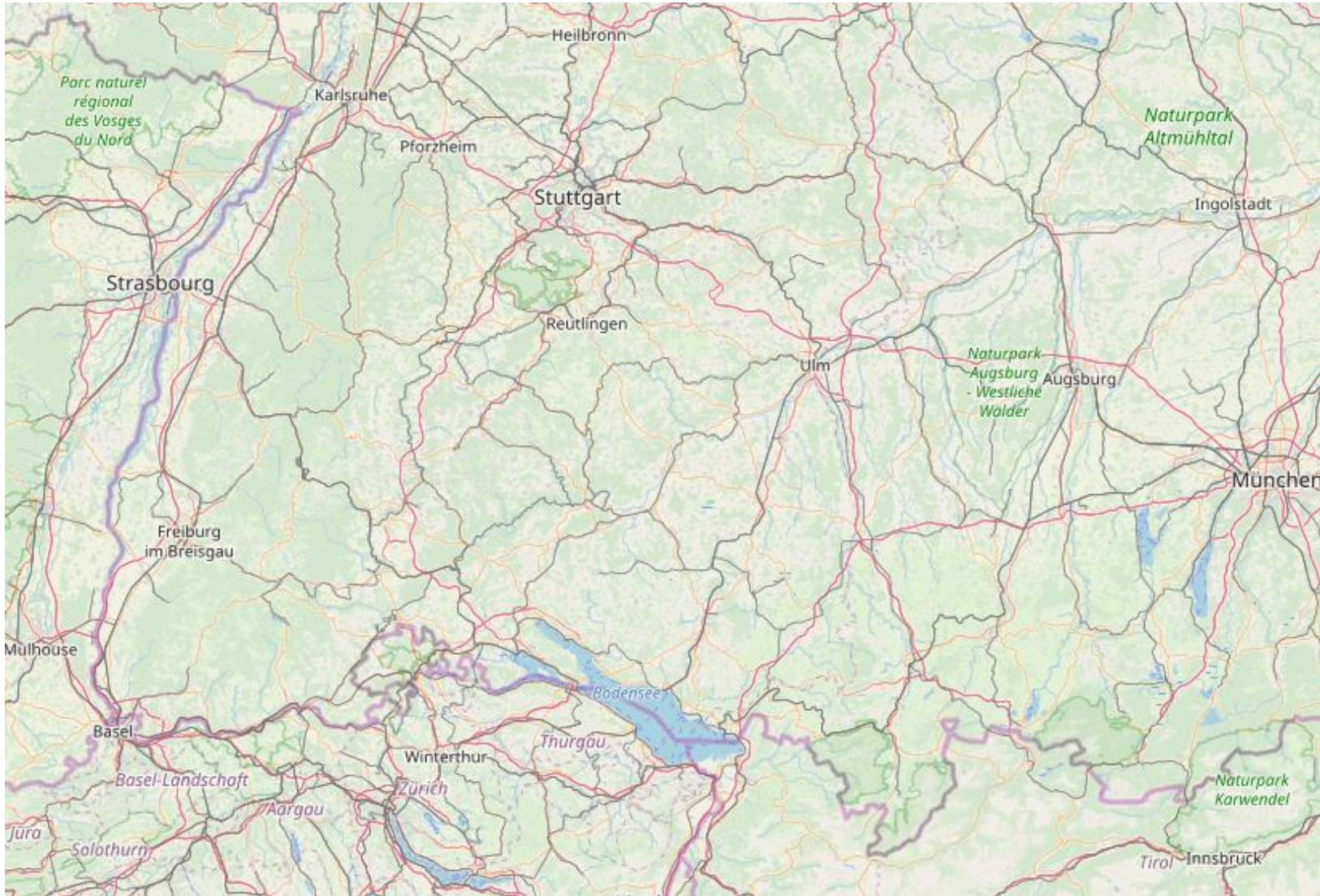
Wo sind die Orte?

Lauchingen

Waldshut-
Tiengen

Titisee-
Neustadt

Gundel-
fingen



Kartendaten © OpenStreetMap-Mitwirkende

Wo sind die Orte?

```
▼ tweet: Status(_api=<tweepy.api.API object at 0x000001C2F97346D8>, _json={'created_at': 'Fri Feb 14 10:00:30 +0000 2020', 'id': 1228257701064531969, 'id_str': '1228257701064531969', 'text'  
> author: User(_api=<tweepy.api.API object at 0x000001C2F97346D8>, _json={'id': 847056191389872128, 'id_str': '847056191389872128', 'name': 'Polizei Freiburg', 'screen_name': 'PolizeiFR',  
  contributors: None  
  coordinates: None  
> created_at: datetime.datetime(2020, 2, 14, 10, 0, 30)  
> entities: {'hashtags': [{...}, {...}], 'symbols': [], 'urls': [{...}], 'user_mentions': []}  
  favorite_count: 0  
  favorited: False  
  geo: None  
  id: 1228257701064531969  
  id_str: '1228257701064531969'  
  in_reply_to_screen_name: None  
  in_reply_to_status_id: None  
  in_reply_to_status_id_str: None  
  in_reply_to_user_id: None  
  in_reply_to_user_id_str: None  
  is_quote_status: False  
  lang: 'de'  
  place: None  
  possibly_sensitive: False  
  retweet_count: 2  
  retweeted: False  
  source: 'SocialHub by maloon'  
  source_url: 'https://app.socialhub.io/'  
  text: '👤 🗨️ #Gundelfingen: Geisterfahrer auf #B3 zwischen #Emmendingen und #Freiburg - Die #Polizei sucht Zeugen:... https://t.co/ZnWJdwblEE'
```

coordinates: None

place: None

Zielsetzung

- Koordinaten der Ortsangaben aus dem Tweet herausfinden und darstellen
- Prototyp zur räumlichen Verortung von Tweets
 - in Deutscher Sprache
 - von Polizeibehörden
 - basierend auf existierenden Werkzeugen

Inhaltsverzeichnis

Teil 1: Methodik

Teil 2: Prototyp

Teil 3: Evaluation

Inhaltsverzeichnis

Teil 1: Methodik

Teil 2: Prototyp

Teil 3: Evaluation

Recherche

Location Name Extractor

- Für die Auswertung von englischsprachigen Tweets in drei Überflutungsregionen programmiert und trainiert

Mordecai

- Für die Auswertung von englischsprachigen Tweets zur Ermittlung von Koordinaten programmiert

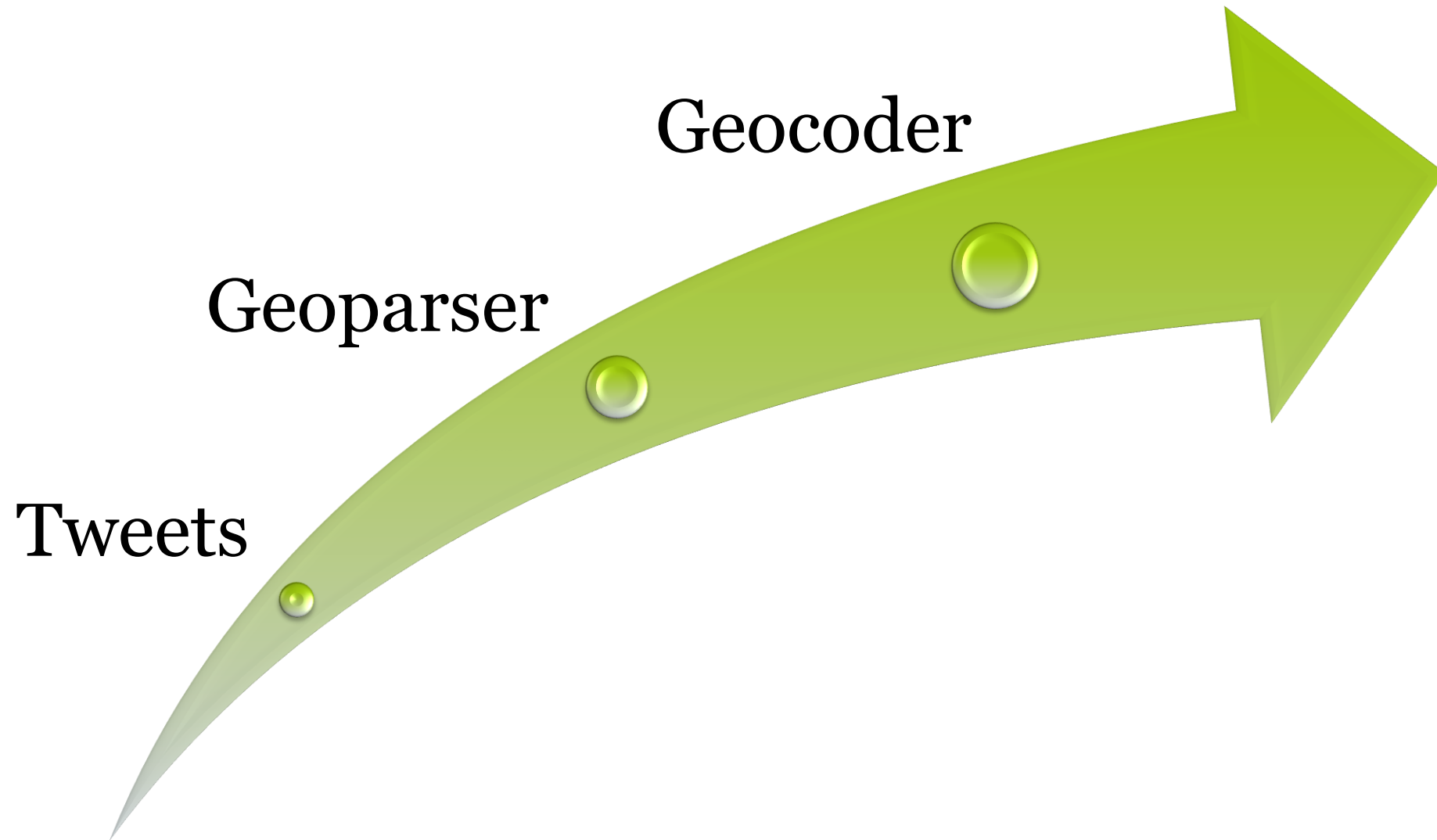
→ Bestehende Programme für die englische Sprache vorhanden

Al-Olimat, H. S., Thirunarayan, K., Shalin, V., & Sheth, A. (29. Juli 2019). *Location Name Extractor*. Von GitHub: <https://github.com/halolimat/LNEx> abgerufen



Halterman, A. (08. August 2019). *mordecai - Full text geoparsing*. Von GitHub: <https://github.com/openeventdata/mordecai> abgerufen

Bestandteile



Herausforderungen



Tweets

- Kurz, Sonderzeichen
- Verlinkungen
- Bilder/Videos
- Keine vollständigen Sätze

Ortsangaben

- Ortsnamen kommen mehrfach vor
- Ortsnamen können abgekürzt sein
- Rechtschreibung nicht immer richtig

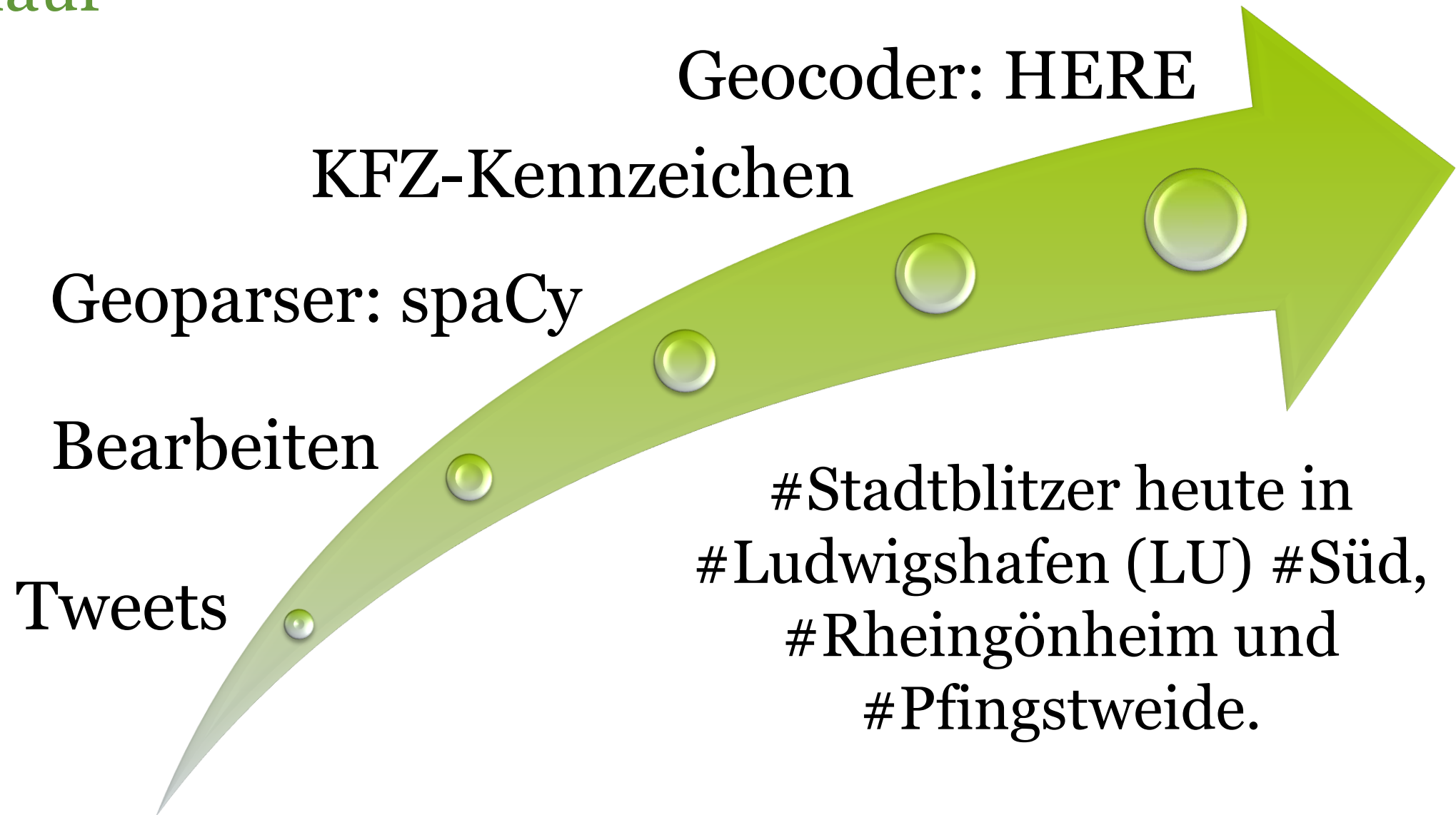
Inhaltsverzeichnis

Teil 1: Methodik

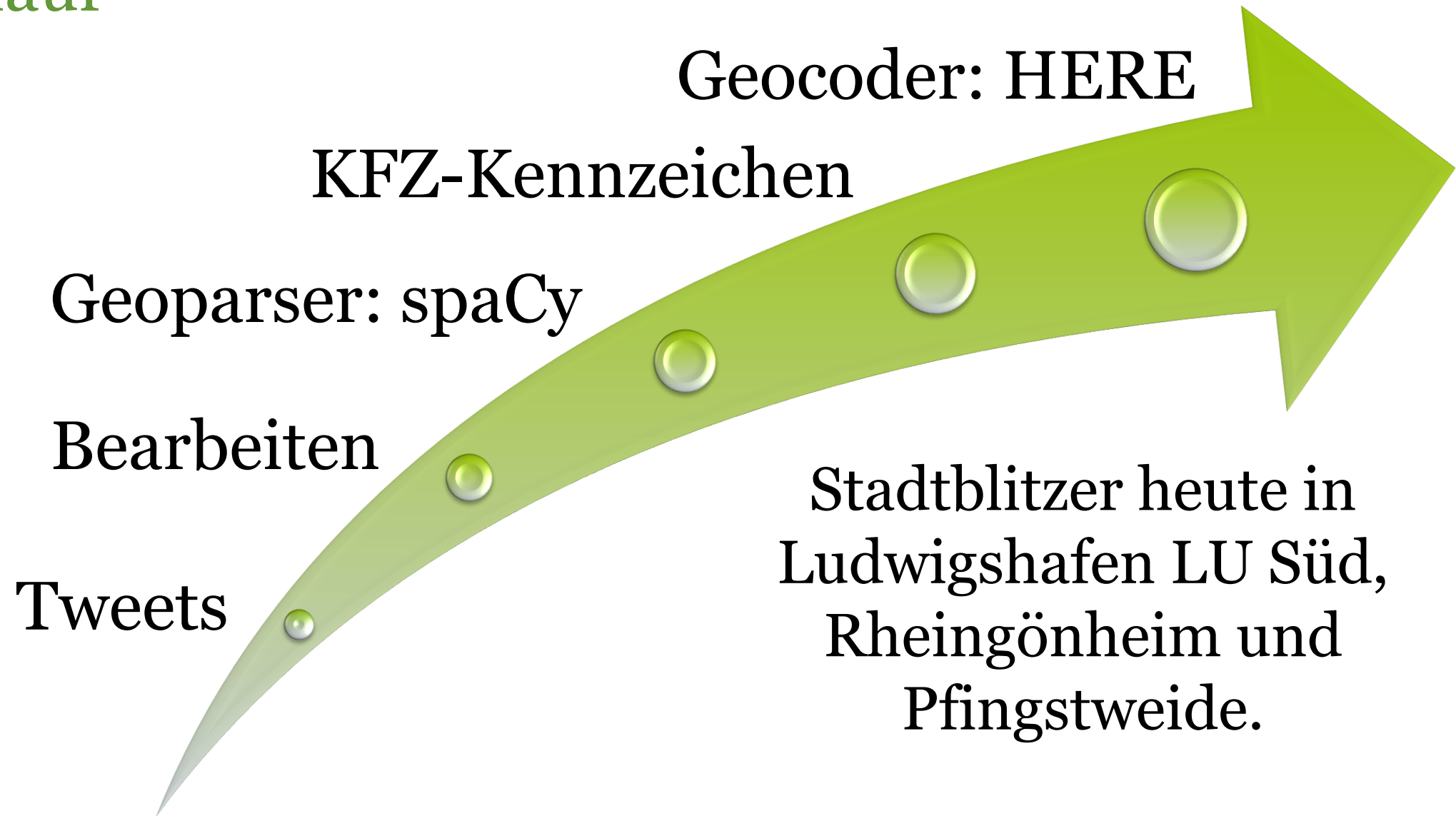
Teil 2: Prototyp

Teil 3: Evaluation

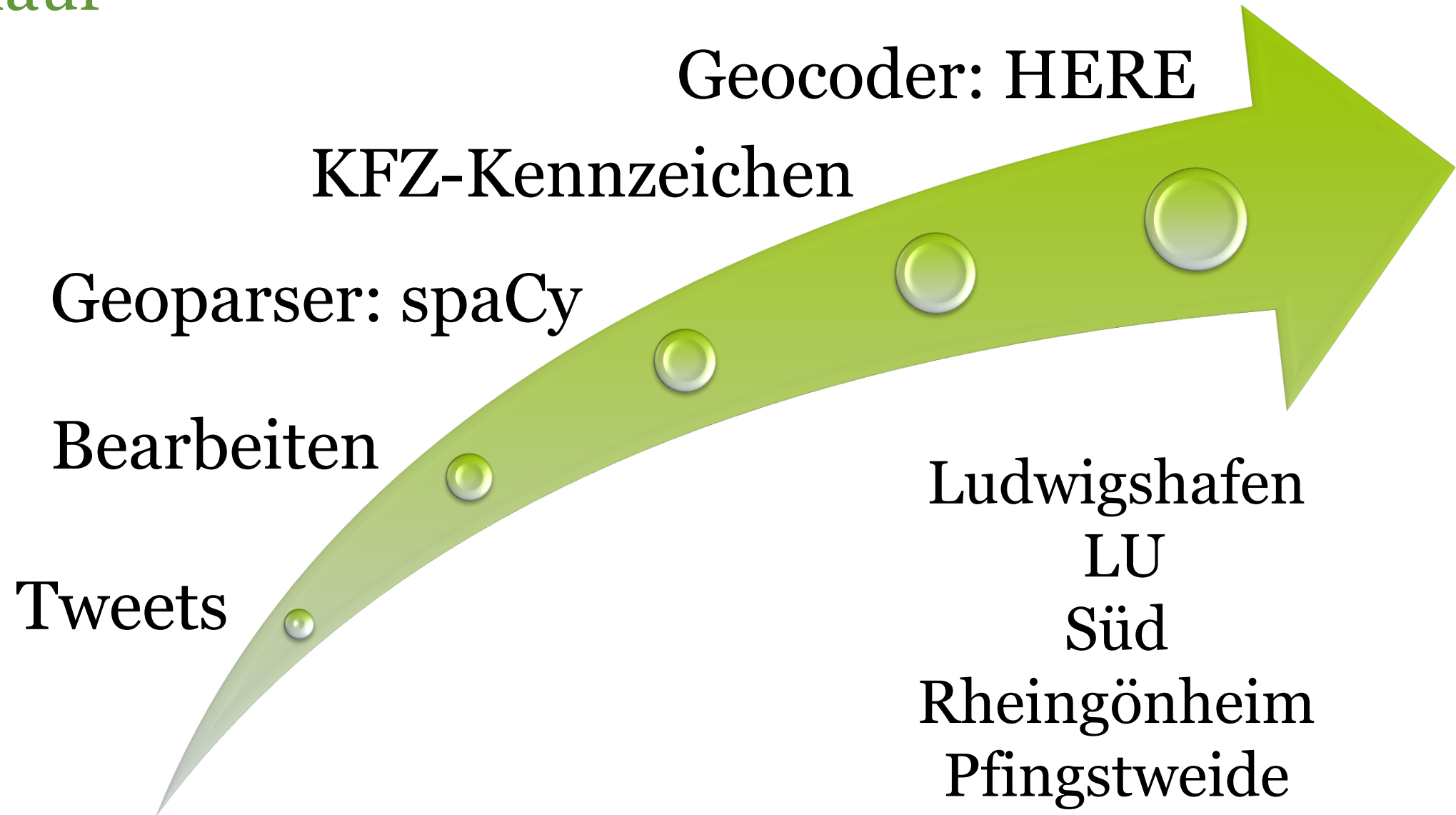
Ablauf



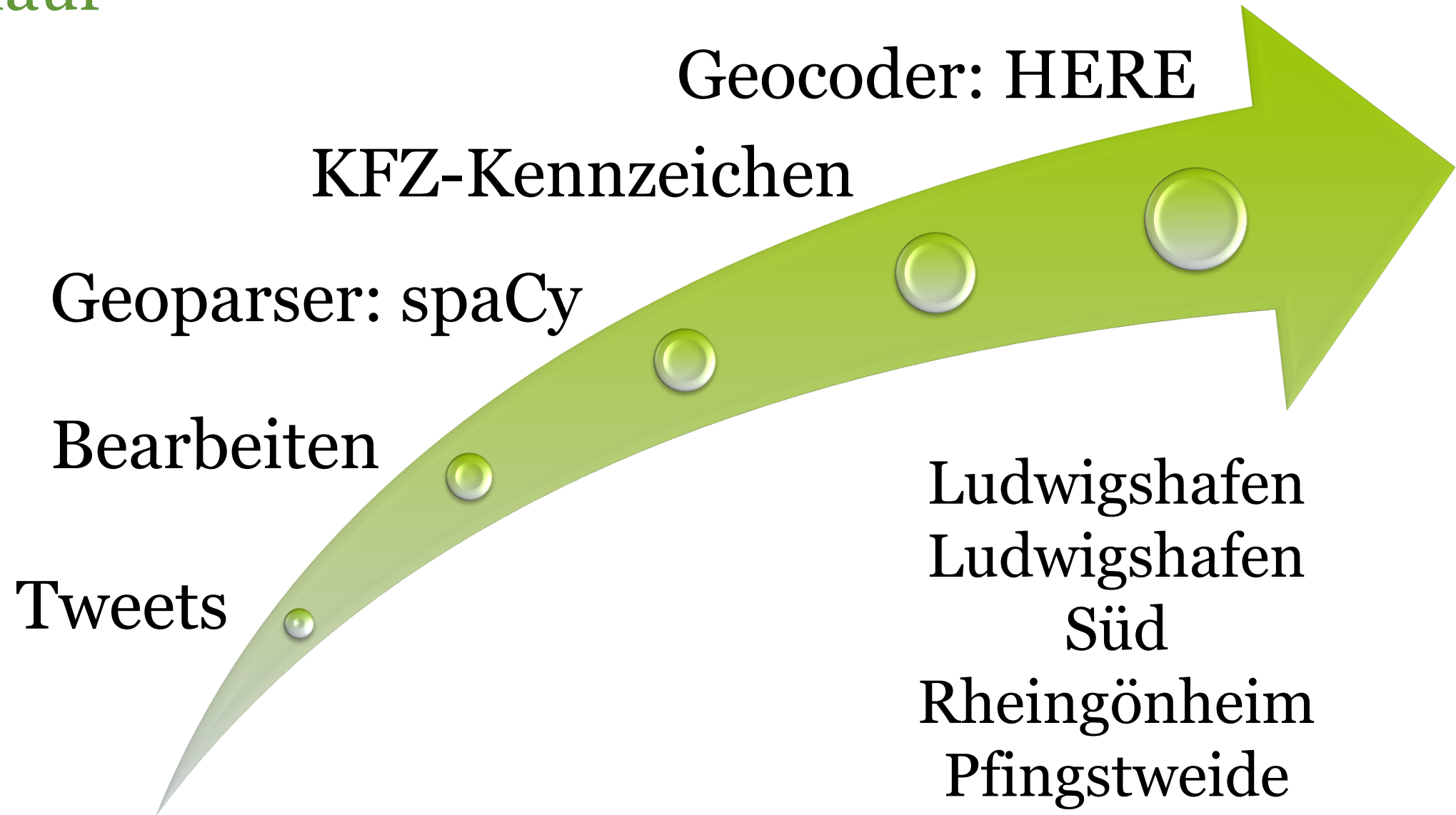
Ablauf



Ablauf



Ablauf



Ablauf



Resultat

The screenshot shows a web browser window with a single tab titled "Bachelorarbeit". The address bar shows "localhost/code/startseite.html". The browser's taskbar includes various icons and a search bar with the text "Zur Suche Text hier eingeben".

The main content area displays a presentation slide with a green background. At the top, it reads "Technik Hochschule Mainz University of Applied Sciences". The title of the slide is "Räumliche Verortung von textbasierten Social-Media-Einträgen am Beispiel von Polizei-Tweets" by "VON SVENJA RUTHMANN, ALEXANDER ROLWES, KLAUS BÖHM".

Below the slide, there is a section titled "EINSTELLUNGEN ZUR AUSWERTUNG DER TWEETS". It contains the following text and options:

Die Tweets der folgenden Polizeibehörde sollen verwendet werden:

- Polizei Brandenburg
- Polizei Bayern
- Polizei Rheinpfalz
- Polizei Freiburg
- Bundespolizei Baden-Württemberg

Die Anzahl an Orten in einem Tweet kann variieren.
Aufgrund dessen legen Sie hier die maximale Anzahl an Tweets fest, die für die Analyse berücksichtigt werden:

1

To the right of the settings is a map of the Mainz region. A blue location pin is placed on the map near the city of Mainz, specifically in the area of Frankenthal. The map shows various roads, including the A 61 and B 9, and several towns and villages.

Resultat



{"text": "Gundelfingen: Geisterfahrer auf B3 zwischen Emmendingen und Freiburg - Die Polizei sucht Zeugen: WJd WbLEE““,

"date": "2020-02-14 10:00:30",

"author": "Polizei Freiburg",

"location": "B3 Freiburg",

"locLabel": "B3, 79115 Freiburg im Breisgau, Deutschland",

"latitude": 47.98528,

"longitude": 7.8248}

Inhaltsverzeichnis

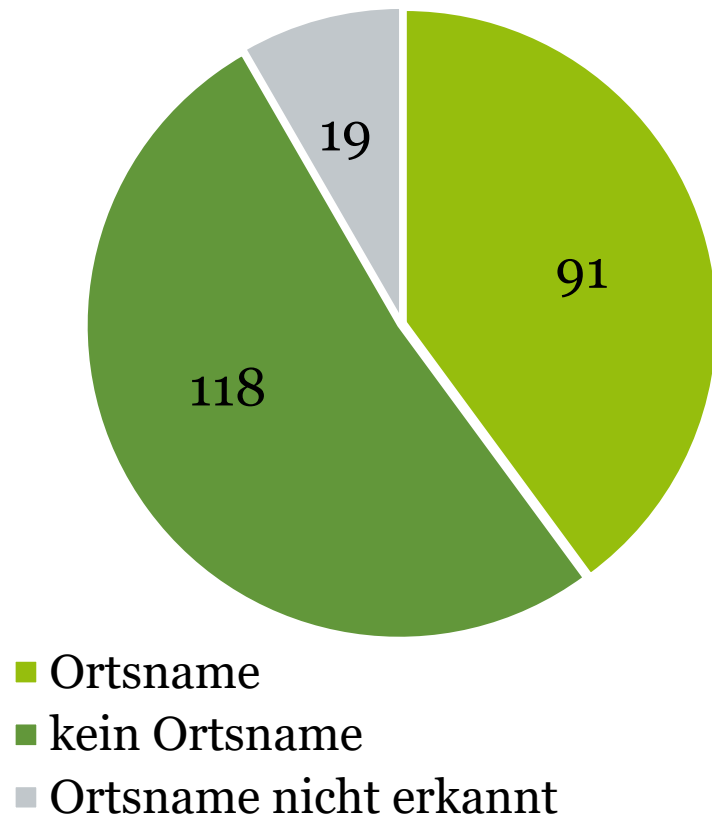
Teil 1: Methodik

Teil 2: Prototyp

Teil 3: Evaluation

Evaluation

Geoparser/spaCy



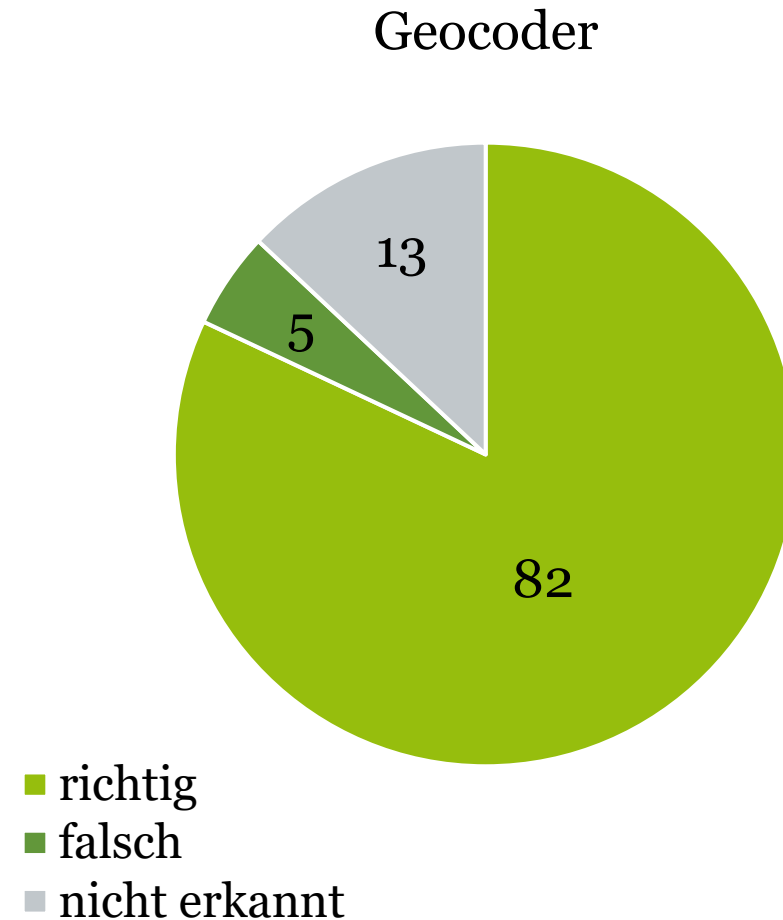
100 Tweets von folgenden
Polizeibehörden wurden getestet

- Polizei Rheinpfalz
- Polizei Brandenburg
- Polizei Bayern
- Polizei Thüringen
- Polizei Sachsen

Evaluation

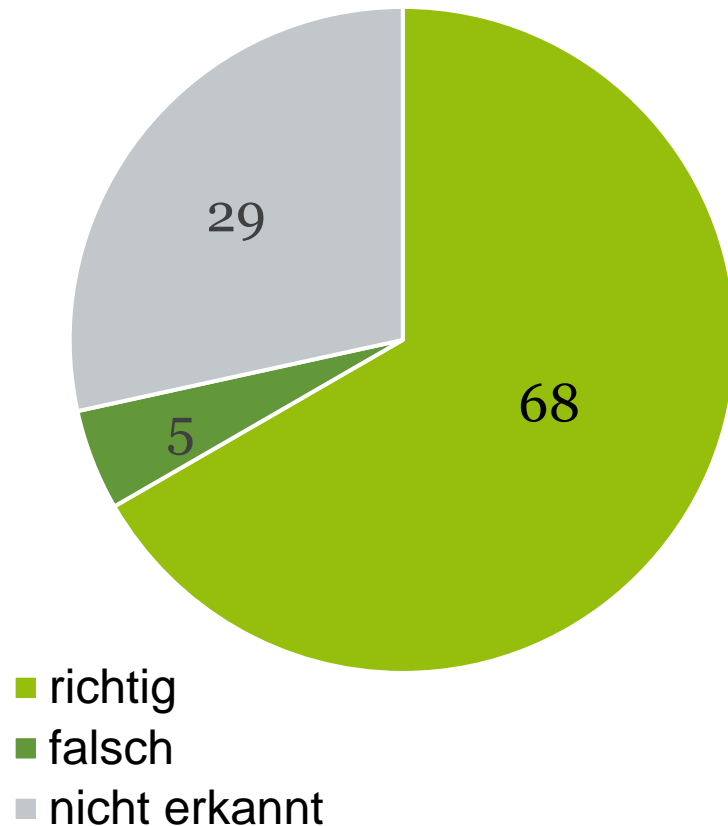
Aus den 100 Tweets wurden

- alle vorhandenen Ortsnamen in die auszulesende Datei geschrieben
- alle sonstigen Eigennamen eliminiert



Evaluation

Gesamtprozess



Gesamtprozess ohne Anpassung

→ 68 richtig erkannte Orte

→ 5 falsch erkannte Orte

→ 29 nicht erkannte Orte

Fehlerquellen

Geoparser

- Erkennt Eigennamen, nicht nur Ortsnamen
- Erkennt nicht immer die vollständige Information zum Ort
- Erkennt Orte gar nicht
- Erkennt Ort in Zusammenhang mit anderem Wort

Geocoder

- Nord, Ost, Süd, West, Mitte nicht eindeutig zuordbar
- Kreise und Bundesländer nicht erkannt
- Autobahnennamen ohne zusätzliche Informationen nicht erkannt
- erkennt Orte, die im Tweet keine sind

Verbesserungsmöglichkeiten

Geoparser

- spaCy mit Tweets und Ortsnamen trainieren
- Eigennamen könnten nach Location abgefragt werden
- Vollständiger Ortsname würde erkannt werden
- Eigenschaften eines Tweets können berücksichtigt werden

Geocoder

- Geocoder könnte um Kreise, Bundesländer, Autobahnen erweitert werden

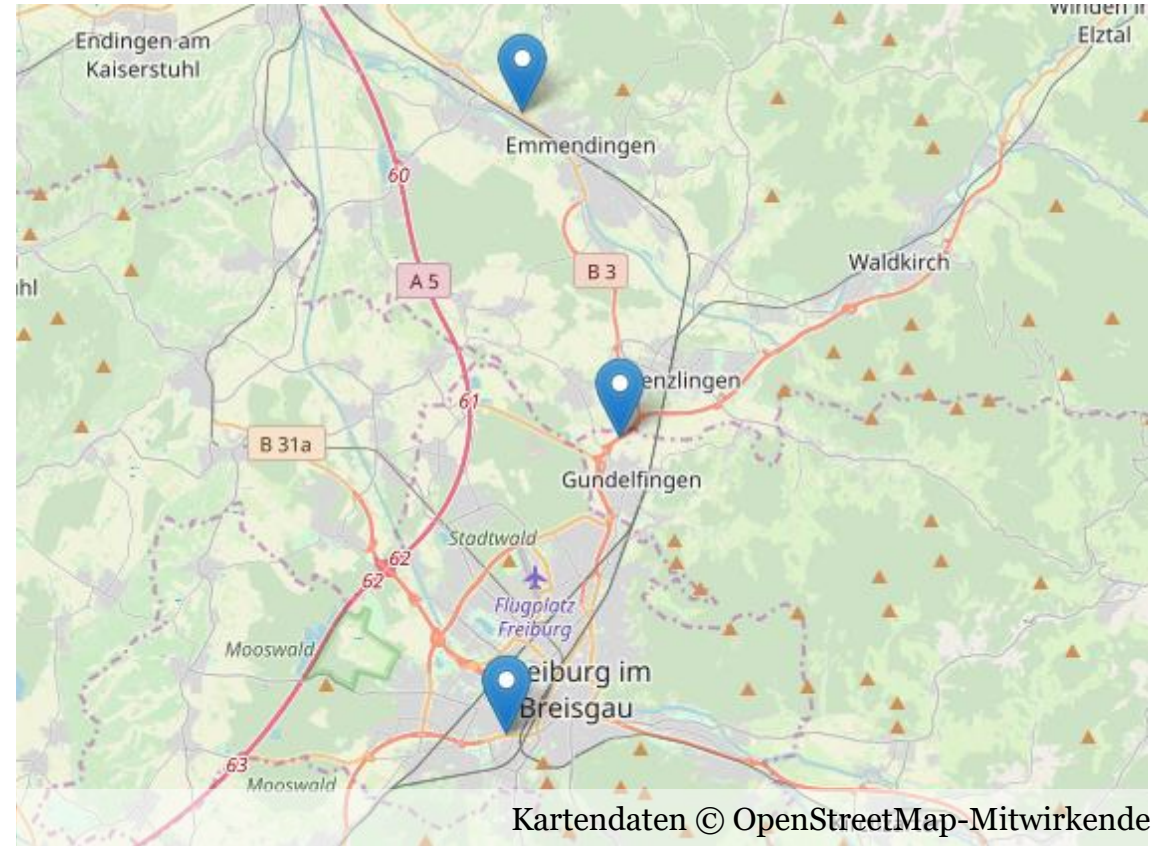
Vielen Dank für Ihre
Aufmerksamkeit



Polizei Freiburg @PolizeiFR · 1 Std.

#Gundelfingen: Geisterfahrer auf #B3 zwischen #Emmendingen und #Freiburg - Die #Polizei sucht Zeugen: presseportal.de/blaulicht/pm/1...

Zeugen werden gebeten sich beim Polizeirevier Freiburg Nord ☎️ 0761 882-4221 zu melden.



Kartendaten © OpenStreetMap-Mitwirkende