Clustertreffen des CC „Authority Files and Community-driven Vocabularies“

Historische Geodaten in den Briefen an Goethe: Ergebnisse einer Masterarbeit

am **11.11.2025**

Marthe Küster

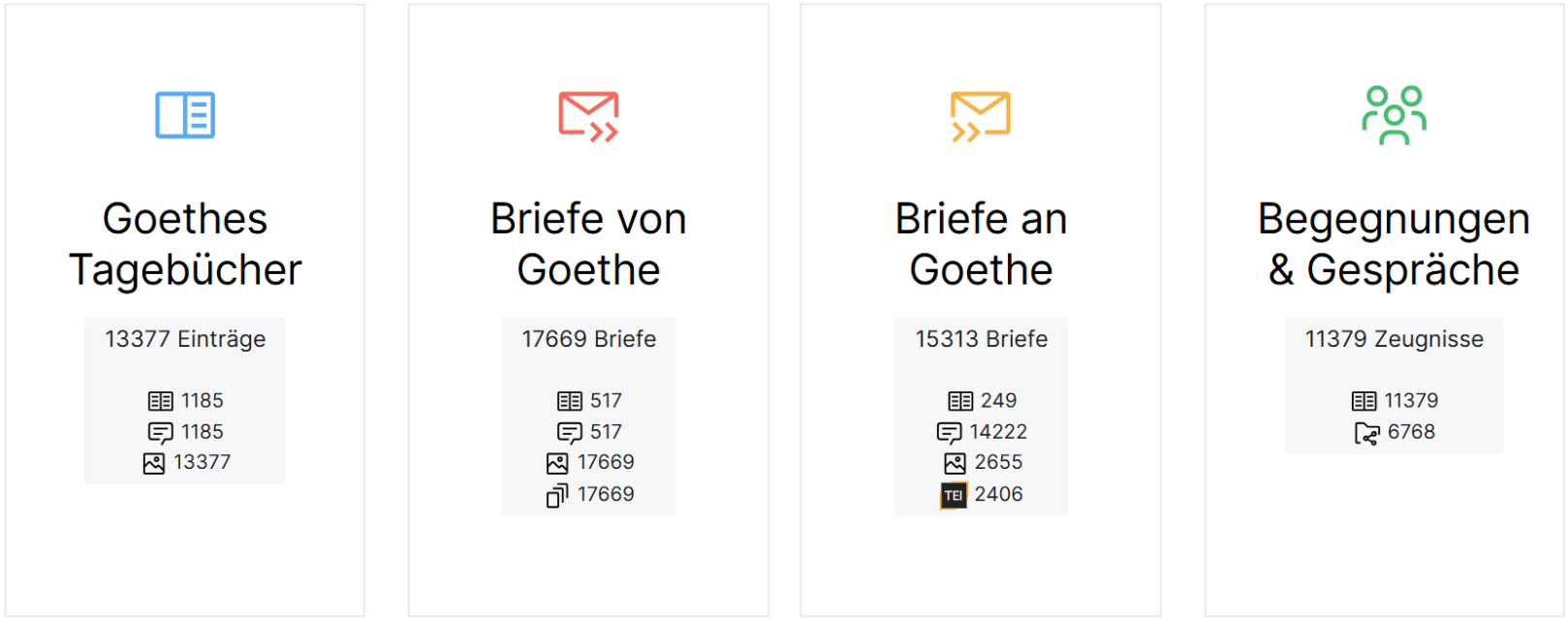
Vorstellung

- 2019-2025: Studium am Institut für Bibliotheks- und Informationswissenschaft an der Humboldt-Universität zu Berlin
- 2022-2025: Studentische Hilfskraft bei TELOTA an der BBAW

Einleitung

- Thema: Die Welt in Briefen: Herausforderungen bei der Erschließung von historischen Geodaten in Digitalen Editionen
- Datengrundlage: Die Briefe an Goethe, die von der Klassik Stiftung Weimar über die Forschungsplattform PROPYLÄEN - Goethes Biographica schrittweise bereitgestellt werden

Einleitung



(Klassik Stiftung Weimar, Sächsische Akademie der Wissenschaften zu Leipzig & Akademie der Wissenschaften und der Literatur Mainz, 2024)

Forschungsstand

Digitale Edition

- Digitale Editionen ermöglichen im Gegensatz zu Druckeditionen
  - Kostengünstige Einbindung von Transkriptionen und Abbildungen
  - Technische Analysemöglichkeiten (Hyperlinks oder Volltextsuche)
  - Schrittweise Veröffentlichung der Daten

  (Sahle, 2013, S. 57, S. 132, S. 148-155)

Der Ortsbegriff

Geografika,
Ortsangabe,
Ortsdaten,
Geodaten

GIS vs. Gazetteer

Richtlinien und Standards:

- TEI-Richtlinien für die Auszeichnung von XML-Dateien
  - Basisformat des Deutschen Textarchivs
- Regeln für die Schlagwortkatalogisierung (RSWK)
- Bibliothekarischen Standard RDA
  - RDA DACH des Standardisierungsausschusses (STA) der DNB
- CIDOC CRM für Museen und ISAD (G) für Archive
- Dublin Core als einfacher Metadatenstandard
- Minimaldatensatz für historische Orte im deutschsprachigen Raum

Entitätscodierung der GND

Code	Definition
g	
gib	Bauwerke und Bauensembles, Monumentalplastiken, Denkmäler, Grabmäler u.ä. gemäß RSWK §730 + 731
gif;gik	Gliedstaaten (sie werden immer mit "gik" doppelt codiert)
gik	Gebietskörperschaften und Verwaltungseinheiten
gil;gik	Länder, Staaten (selbstständig) (sie werden immer mit "gik" doppelt codiert)
gin	Natürliche geografische Einheiten, biogeografische und paläografische Einheiten, Namen von geografischen Teileinheiten mit Himmelsrichtung bzw. mit anderen Wendungen gemäß RSWK § 205,1 und 2
gio	Kleinräumige Geografika innerhalb eines Ortes, ausgenommen Bauwerke und Bauensembles
gir;gik	Geistliche Reichsfürstentümer (bis 1803), Verwaltungseinheiten der Ostkirche (sie werden immer mit "gik" doppelt codiert)
giv;gik	Verwaltungseinheiten, die gemäß der EH-G-03 einen instantiellen Oberbegriff von dem Gattungsbegriff der Verwaltungseinheit bekommen (sie werden immer mit "gik" doppelt codiert)
giw	Grenzen, Wege, Linien
gix	Extraterrestrika
giz	Alle geografischen Namen, die sich keinem der anderen Entitäten-Untertypen zuordnen lassen, Regionen und Gebiete gem. RSWK § 204a, einschließlich des Umlands einzelner Städte und Gemeinden
gxz	Fiktive Orte

Deutsche Nationalbibliothek (Hrsg.). (2021). Entititätscodierung: Vergaberichtlinien. https://wiki.dnb.de/download/attachments/90411323/entitaetenCodes.pdf

Digitale Ortsverzeichnisse

OCLC | VIAF

GeoNames

GND

Verein für Computergenealogie
genealogy.net
compgen.de

Das Geschichtliche Orts-Verzeichnis

Ortsname:

Suche

bavarikon

ortsnamen.ch
Das Portal der schweizerischen Ortsnamenforschung

PLEIADES

ISGV

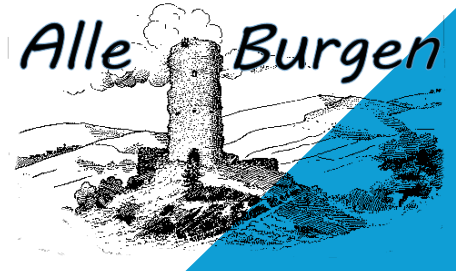
HISTORISCHES ORTSVERZEICHNIS VON SACHSEN

leo bw
Landeskunde entdecken online

HISTORISCHES LEXIKON der SCHWEIZ HLS

WIKIDATA

Alle Burgen



FactGrid
a database for historians.

Landesgeschichtliches Informationssystem Hessen

World Historical Gazetteer
linking knowledge about the past via place

Digitale Ortsverzeichnisse

OCLC | VIAF

GND

Verein für Computergenealogie
genealogy.net
compgen.de

Das Geschichtliche Orts-Verzeichnis

Ortsname:

Suche

Geol

Forschungsdatenbank so:fie

EIADES

bavarikon

ISGV

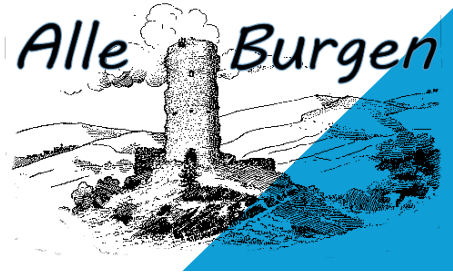
HISTORISCHES ORTSVERZEICHNIS VON SACHSEN

leobw
Landeskunde entdecken online

HISTORISCHES LEXIKON der SCHWEIZ HLS

WIKIDATA

Alle Burgen



Landesgeschichtliches Informationssystem Hessen

HESSEN

FactGrid
a database for historians.

World Historical Gazetteer
linking knowledge about the past via place

Herausforderungen bei der Erfassung von historischen Ortsangaben in Briefen

Bosse (2019):

- Ortsangaben:
  - Genauigkeit der Ortsangaben sind unterschiedlich
  - Sie sind nicht immer mit überliefert oder nur indirekt und unvollständing
- Kartendarstellung
  - Fiktive Orte lassen sich auf einer Karte nicht darstellen
  - Bauwerke, die die Standortkoordinaten selten wechseln lassen sich besser Darstellen als Länder oder Staaten
- Wandel von Orten mit der Zeit durch Eingemeindung, Auflösung, Umbenennung und Verschiebung der Ländergrenzen

Herausforderungen bei der Erfassung von historischen Ortsangaben in Briefen

Schumacher, Flüh & Nantke (2024):

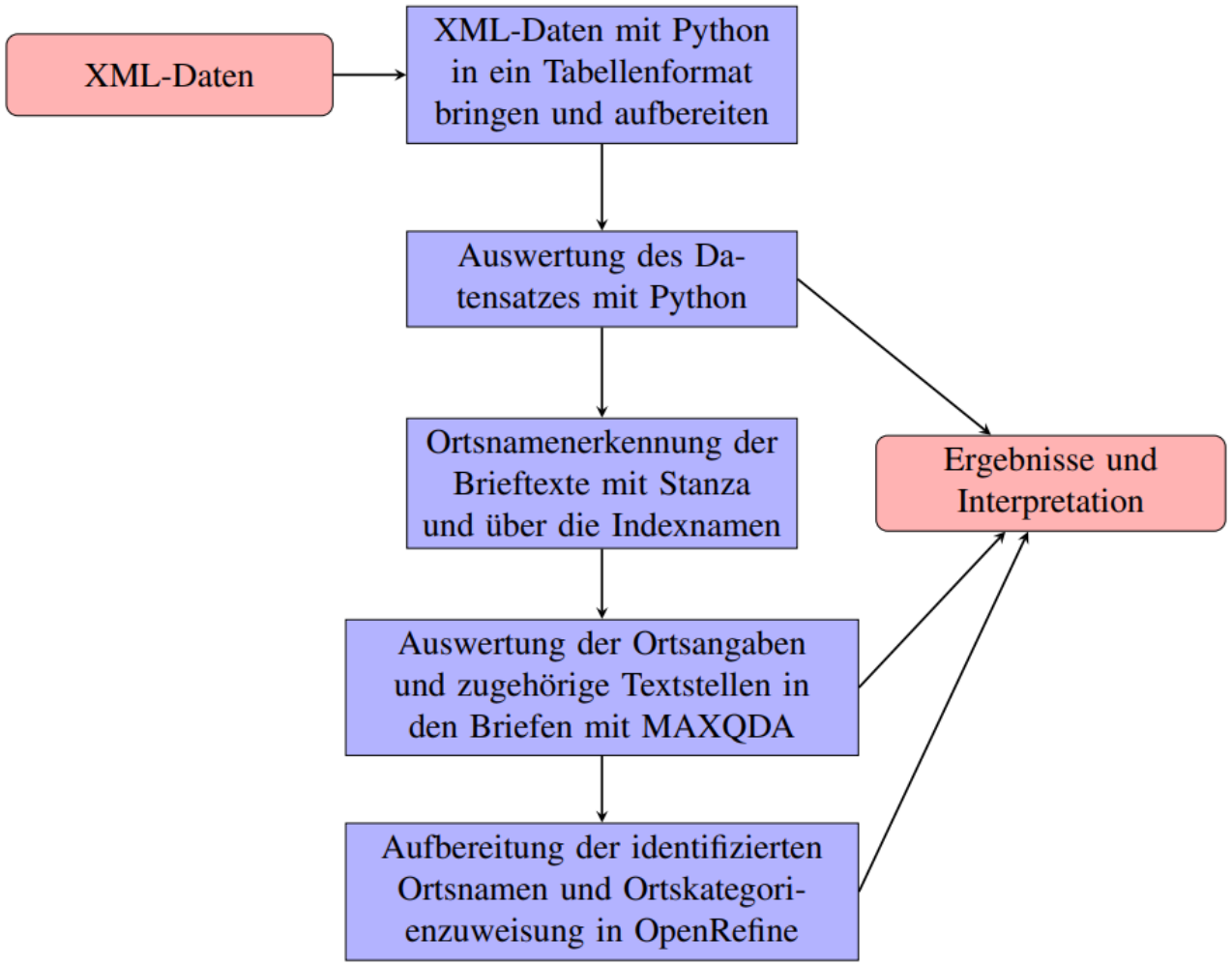
- Aufwendige Workflows bei der händischen Erschließung von Ortsangaben in Brieftexten sind gängige Praxis
- Sie unterscheiden zwischen konkreten Bezeichnungen (z. B. „Berlin“), abstrakte Beschreibungen (z. B. „Vaterland“) und adverbiale Benennungen (z. B. „hier“)

Herausforderungen bei der Erfassung von historischen Ortsangaben in Briefen

Trzeciak (2023):

- untersuchten 30 Liebesbriefe
- unterschieden zwischen erwähnten Orte (’mention’) und Schreiborten der Briefe (’origin’)
- die Erfassung von Personendaten war aufwändiger als die der Ortsangaben, da es sich nicht um Personen des öffentlichen Lebens handelte

Methode


XML-Daten
XML-Daten mit Python in ein Tabellenformat bringen und aufbereiten
Auswertung des Datensatzes mit Python
Ortsnamenerkennung der Brieftexte mit Stanza und über die Indexnamen
Auswertung der Ortsangaben und zugehörige Textstellen in den Briefen mit MAXQDA
Aufbereitung der identifizierten Ortsnamen und Ortskategorienzuweisung in OpenRefine
Ergebnisse und Interpretation

Wichtige Werkzeuge

MAXQDA

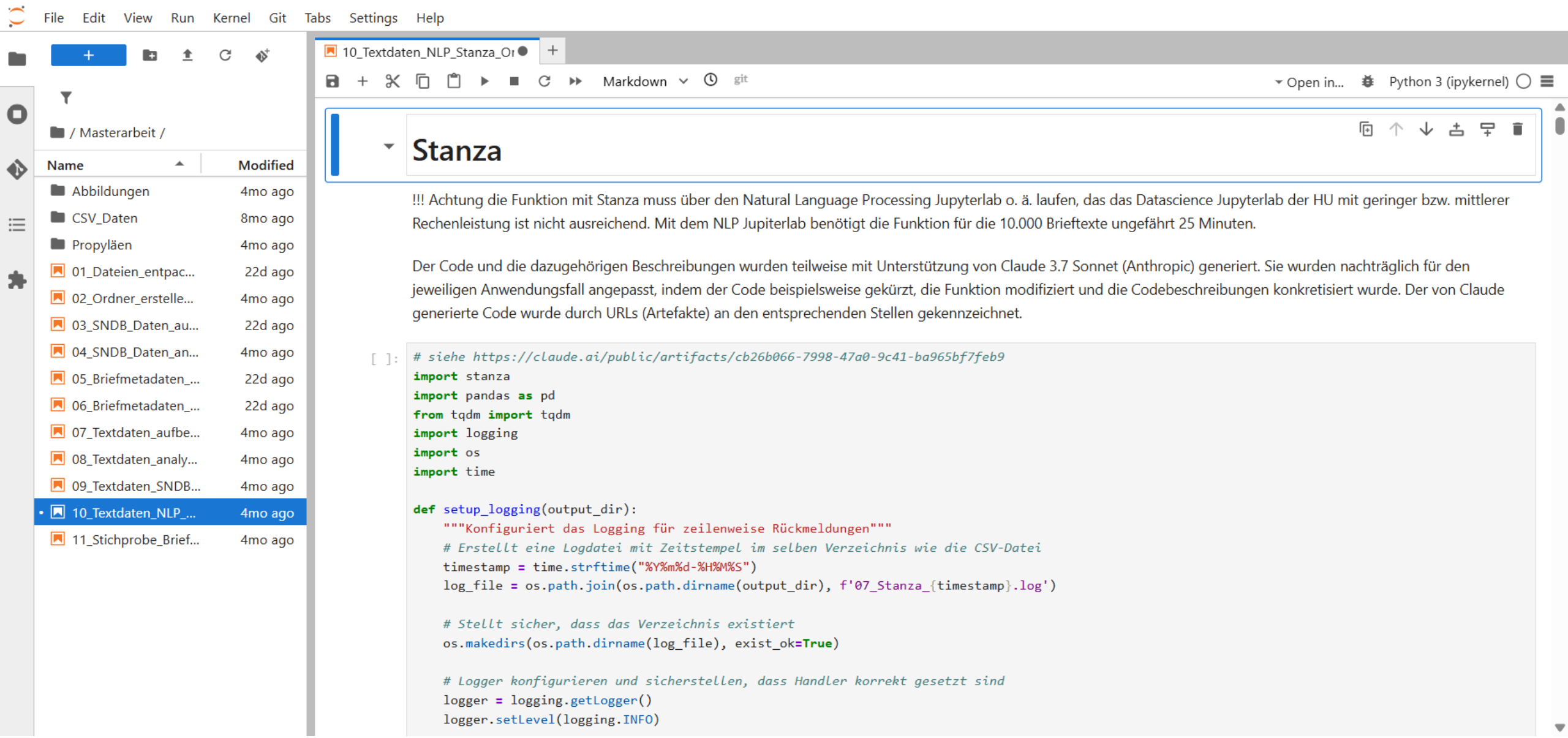
- MAXQDA ist eine Software zur Analyse qualitativer, quantitativer und Mixed-Methods-Daten (vor allem angewendet in den Sozialwissenschaften)

➢lizenzpflichtig

- Texte, Interviews, Umfragen oder audiovisuelle Daten können codiert, kategorisiert und thematisch ausgewertet werden.

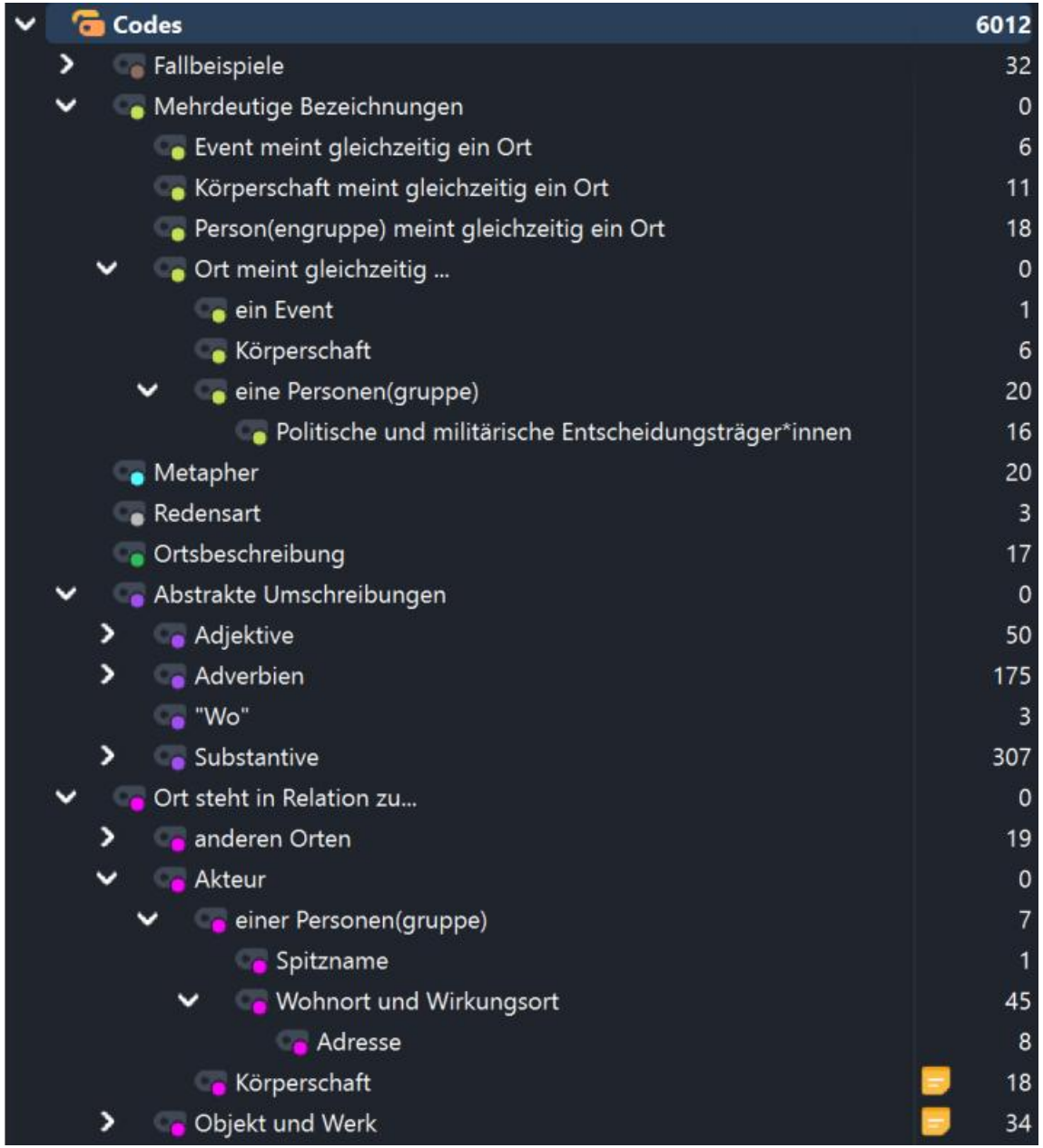
Stanza

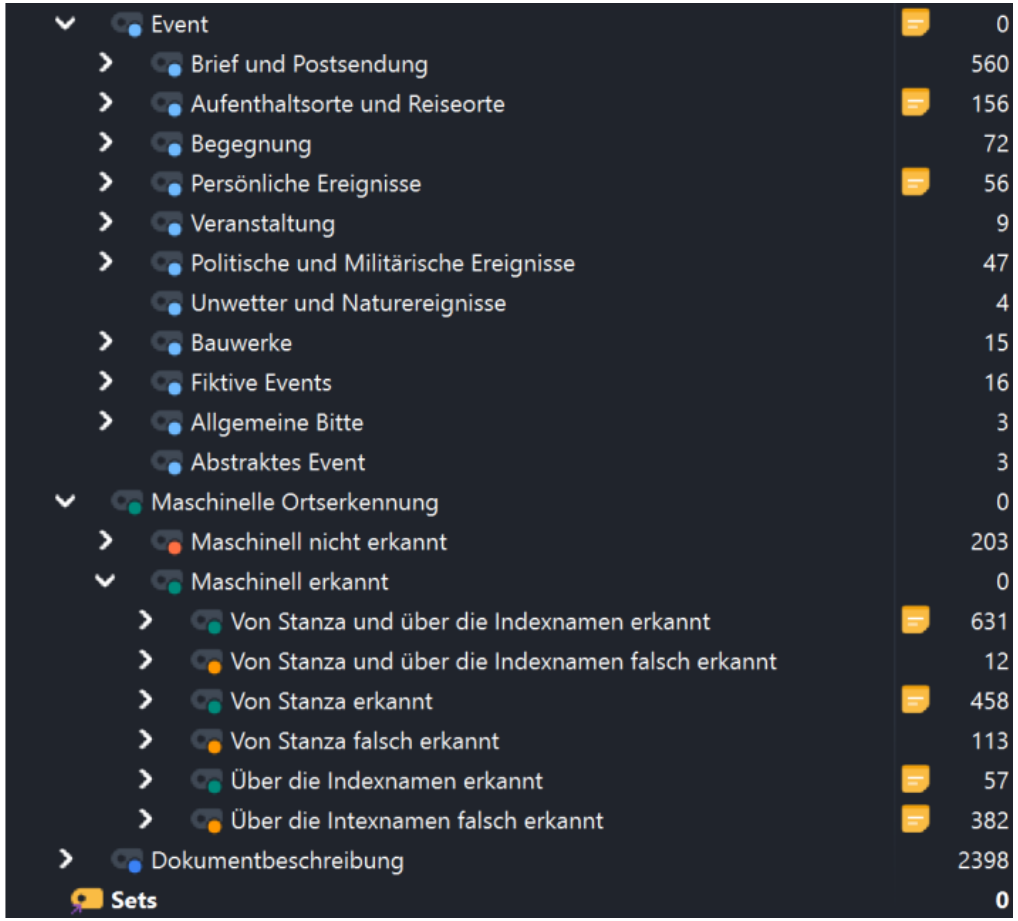
- Von der Stanford University entwickelte Python-Bibliothek für NLP, konkurriert mit CoreNLP, FLAIR und SpaCy
- Kann für Named Entity Recognition genutzt werden
- Kritik an NLPs: Keine Berücksichtigung des Kontextes wie es bei LLMs möglich ist (Hiltmann et al., 2025)



Auswertung des Datensatzes mit Python

MAXDA-Codes

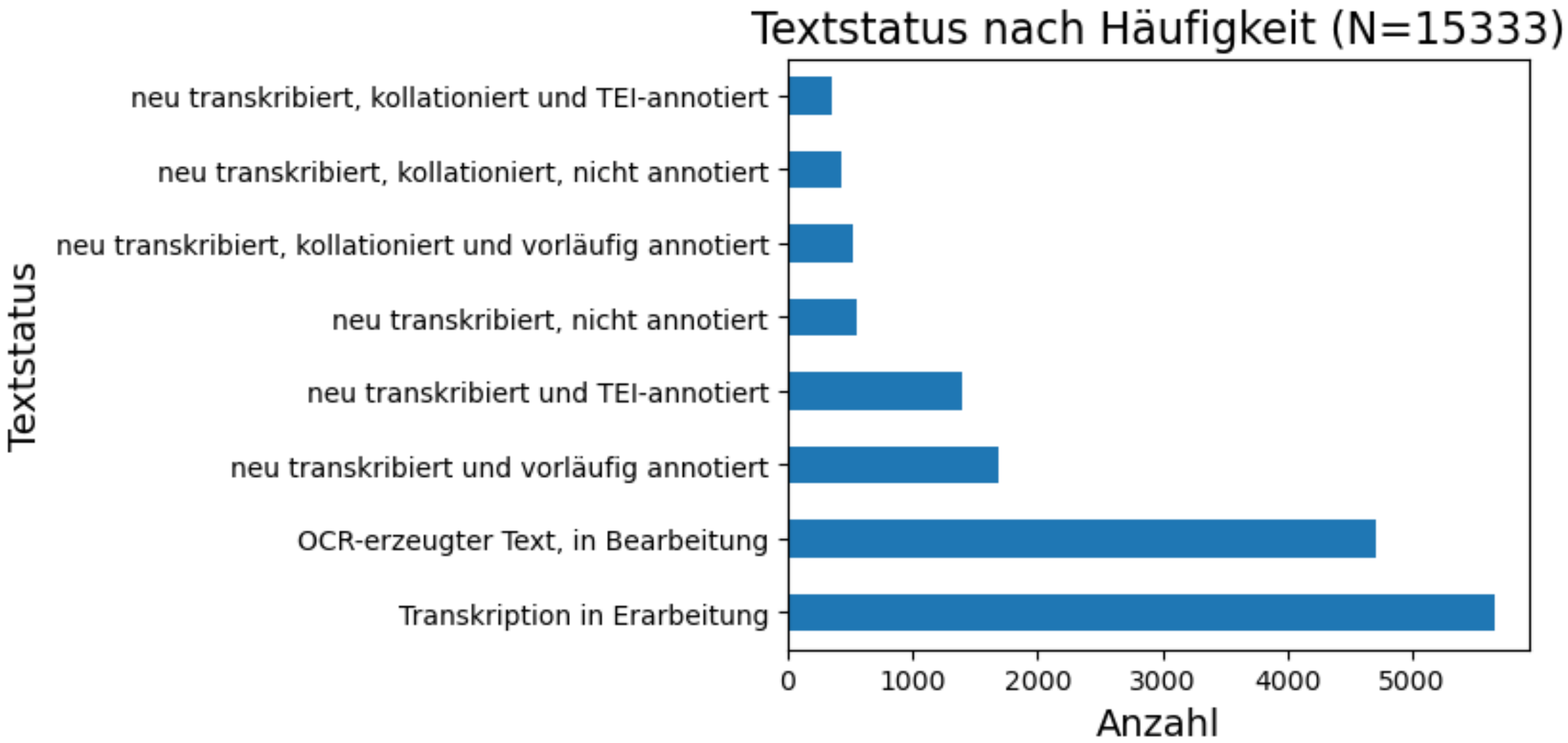


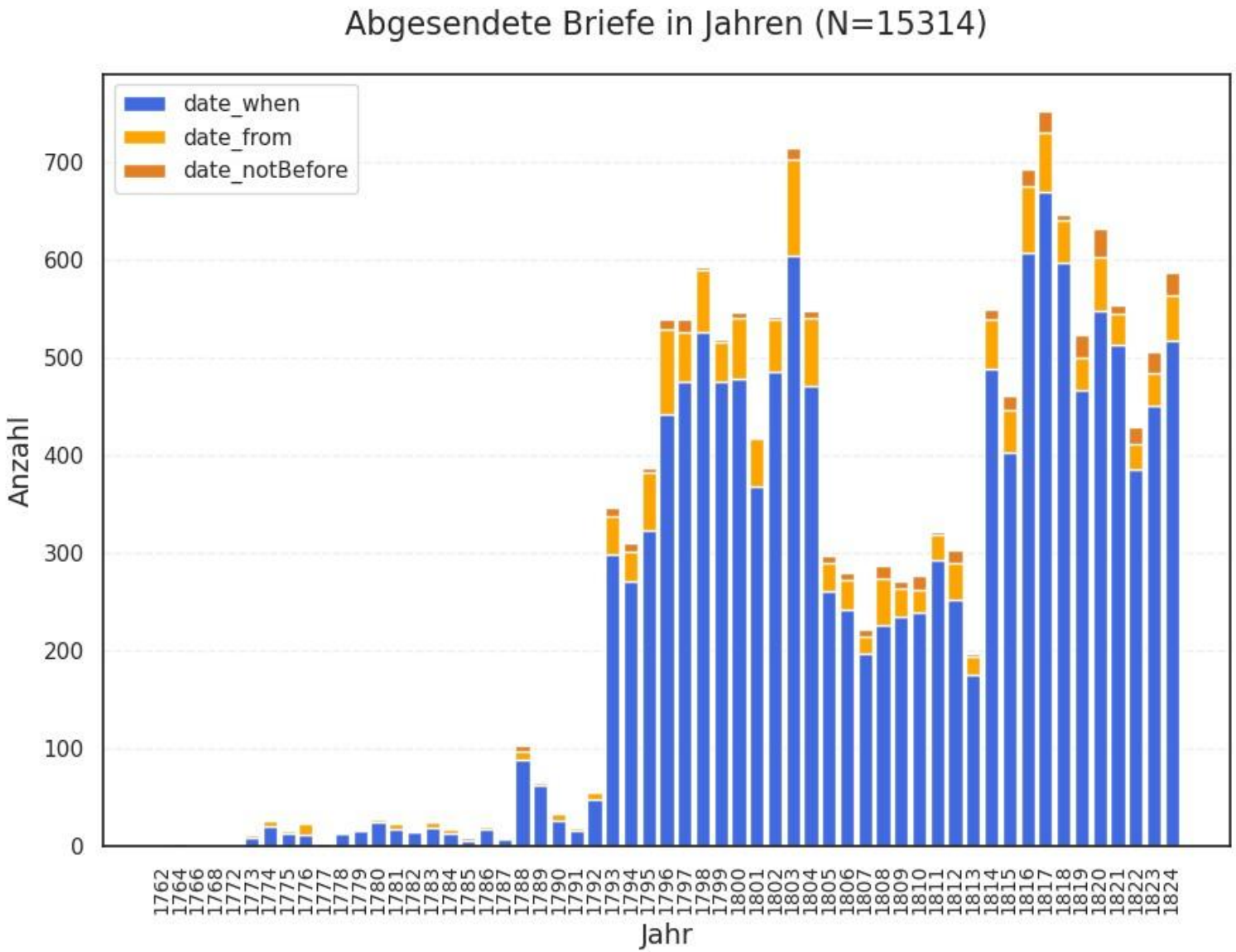


Die zentralen mit MAXQDA erstellten Codes für die Stichprobe des Quellenkorpus Briefe an Goethe.

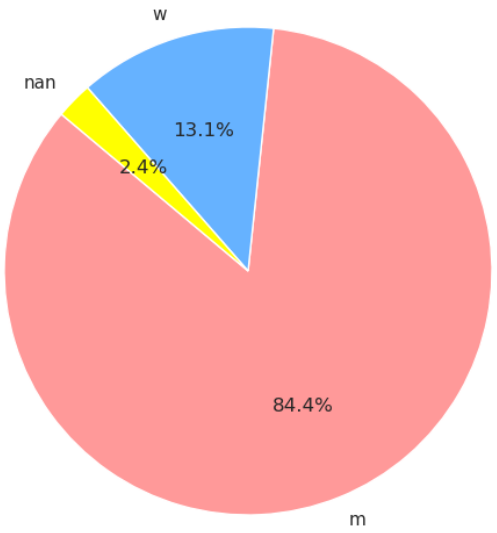
Analyse

- Briefmetadaten:
  - quantitativ mit Python (rund 15.000 Briefe)
- Brieftexte:
  - ausgezeichnete Ortsangaben in den Brieftexten mit Python
  - qualitativ mit MAXQDA (eine Stichprobe von 300 Briefen)

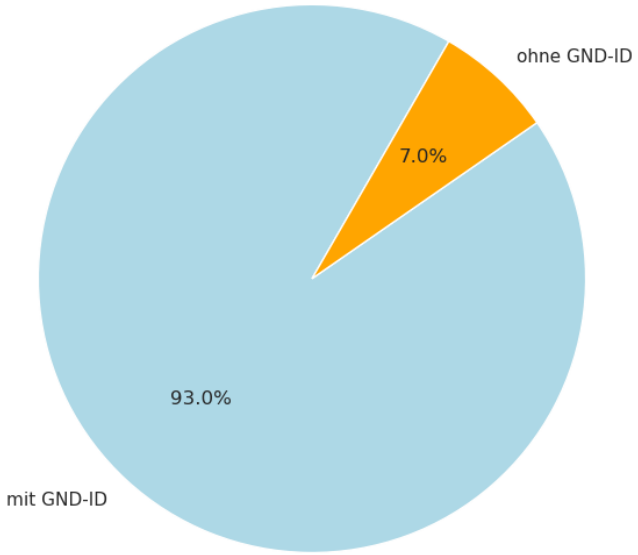

Textstatus nach Häufigkeit (N=15333)
Textstatus
neu transkribiert, kollationiert und TEI-annotiert
neu transkribiert, kollationiert, nicht annotiert
neu transkribiert, kollationiert und vorläufig annotiert
neu transkribiert, nicht annotiert
neu transkribiert und TEI-annotiert
neu transkribiert und vorläufig annotiert
OCR-erzeugter Text, in Bearbeitung
Transkription in Erarbeitung
0
1000
2000
3000
4000
5000
Anzahl


Abgesendete Briefe in Jahren (N=15314)
date_when
date_from
date_notBefore
Anzahl
0
100
200
300
400
500
600
700
Jahr
1762
1764
1766
1768
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824

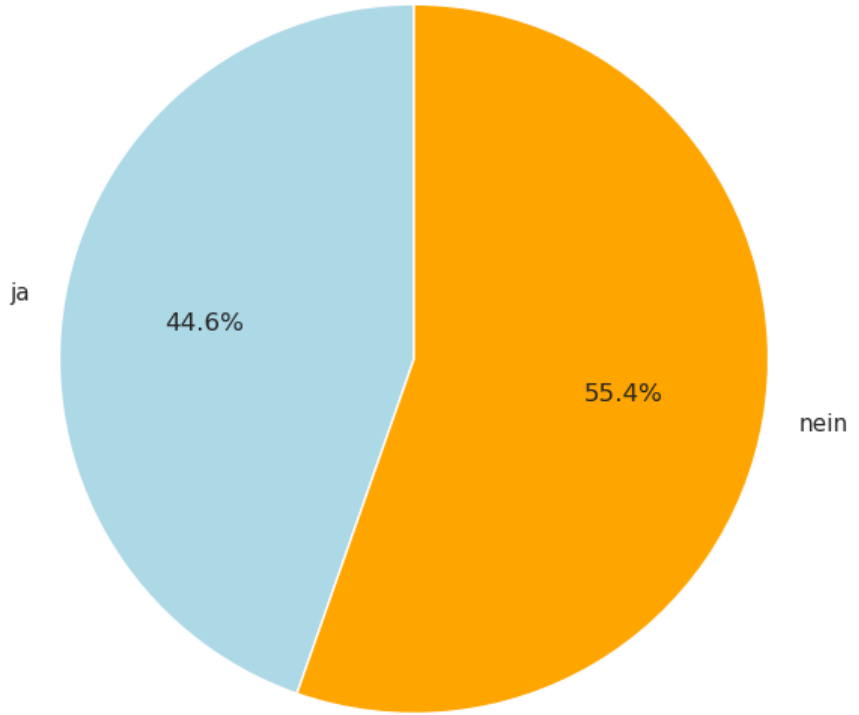
Prozentuale Verteilung Geschlecht der Absender*innen (N=15458)



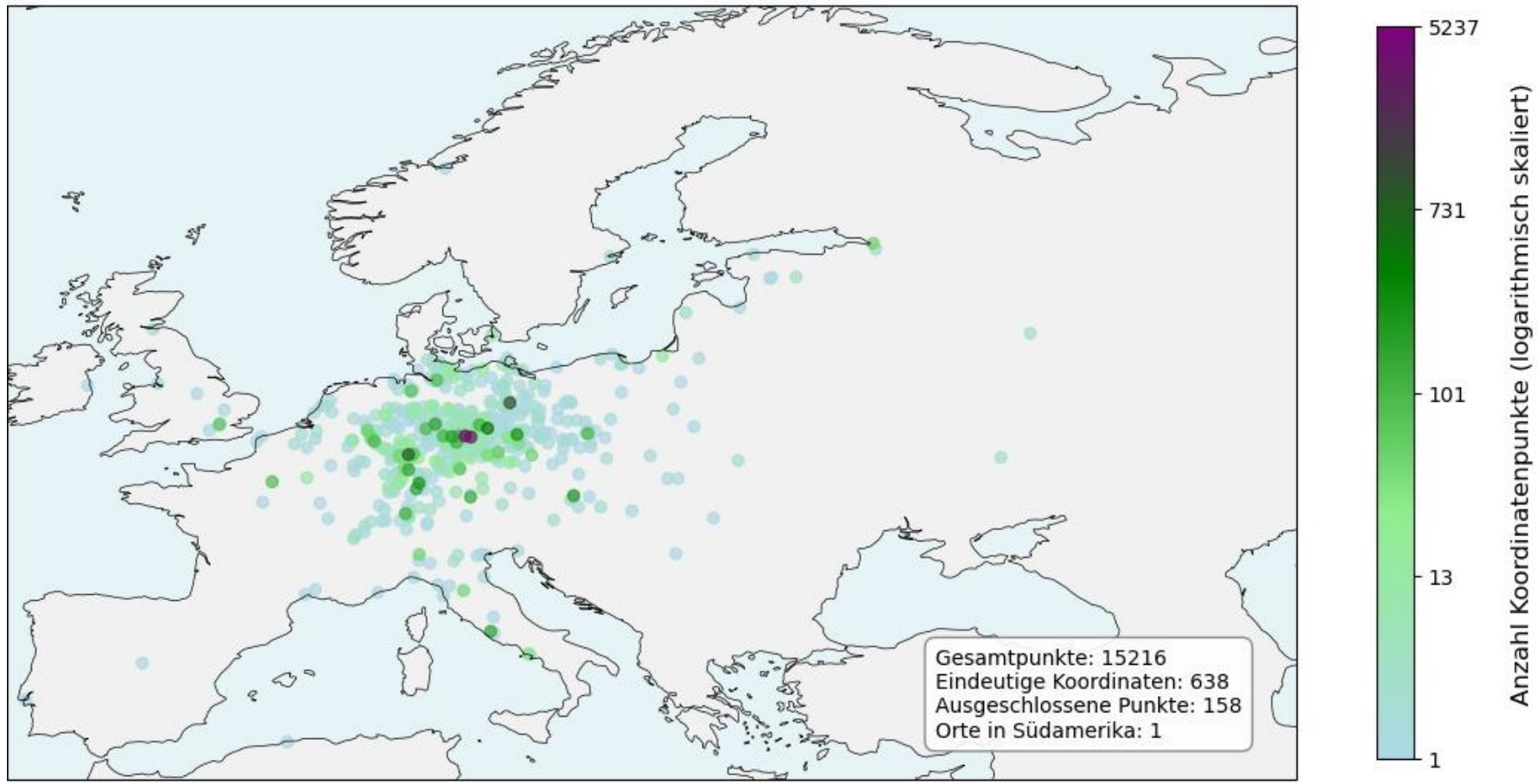
Prozentanteil der vermerkten GND-IDs bei den Absender*innen (N=15458)



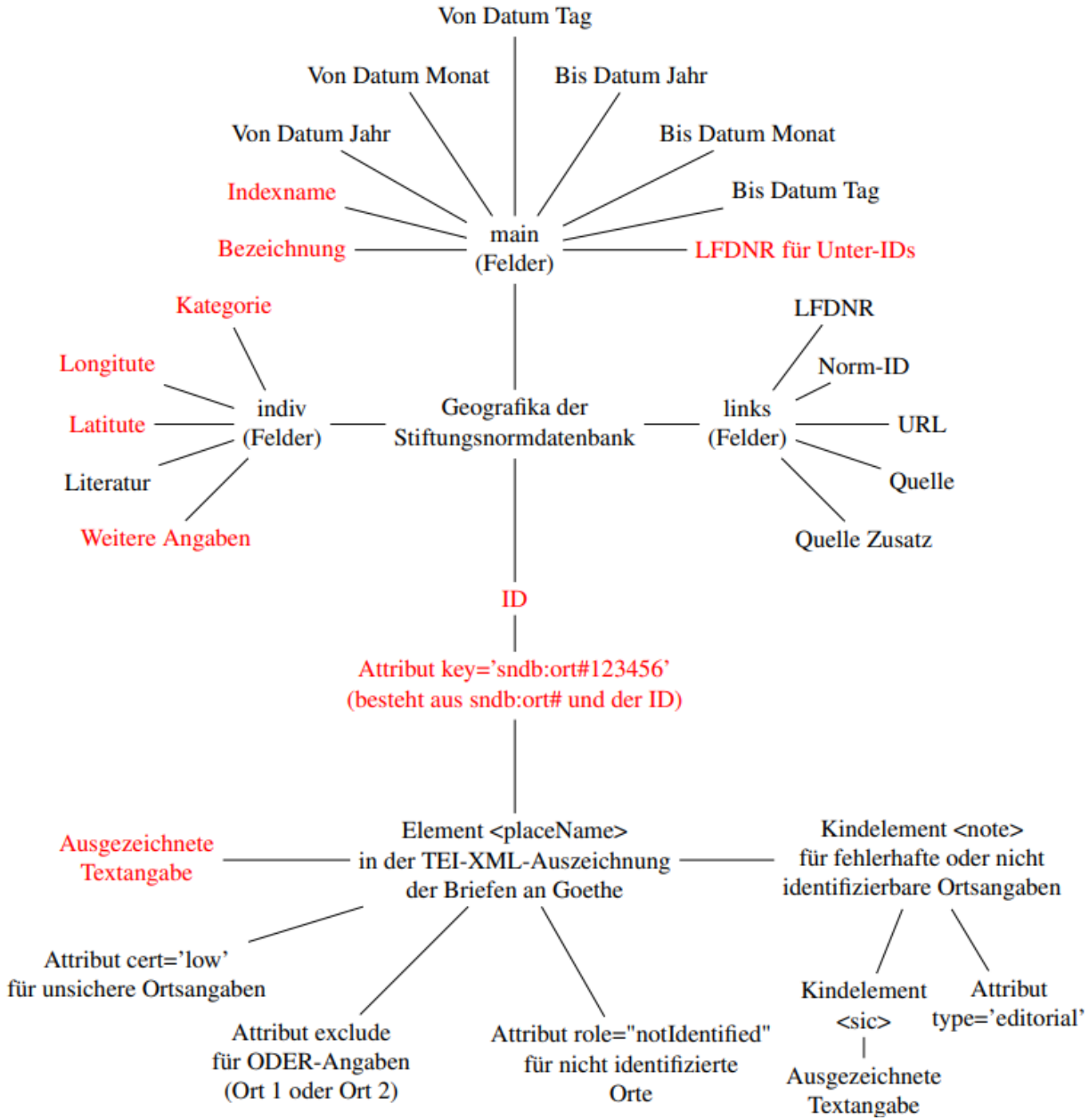
Prozentanteil der Absender*innen mit Adelstitel oder -zusatz (N=15458)



Gesicherte Entstehungsorte der Briefe nach Häufigkeit (N=15051)



Rot markiert:
Pflichtangaben



Bereits ausgezeichnete Ortsangaben im Brieftext

- Rund 36% der Briefe, die bereits TEI-annotiert sind haben keine Ortsangabe im Brieftext
- Die Ortsangaben, die es gibt, sind vielfältig:
  - Beispiel 1: Zum Absendeort „Weimar“ sind in den Brieftexten: unterschiedliche Schreibweisen, unterschiedliche Kürzel, abstrakte Umschreibungen („zu Hause“) und genaue Gebäudeangaben („Im Elephanten“) nachgewiesen
  - Beispiel 2: Für Frankfurt am Main konnten 30 unterschiedliche Schreibweisen und Abkürzungen nachgewiesen werden, ohne dass die Groß- und Kleinschreibung, Kommata oder Doppelpunkte berücksichtigt wurden

Beispiel Schreibweisen Frankfurt am Main

- Schreibweisen Frankfurt ausgeschrieben:
  frankfurt, franckfurt, frankfurth, franckfurth, francfurt, francfort
- Schreibweisen Frankfurt gekürzt: ffurt 20
  Frfurt, ffurth, frankf, ffort, frft, frf, frckfurt, frank, ffrt
- Schreibweisen Frankfurt am Main ausgeschrieben
  franckfurth am main, frankfurt am mayn, am mayn frankfurth, frankfurt am/main
- Schreibweisen am Main gekürzt
  frankfurt a/m, franckfurt a/m, frankfurt a m, frankfurth a/m, frankfurth am, francfort s/m
- Schreibweisen Frankfurt am Main gekürzt
  ffurt a/m, frankf a/m, ff a/m, f a/main

Unterschiedliche Schreibweisen insgesamt = 30*

* Groß und Kleinschreibung, Leerzeichen davor und danach, Zeilenumbrüche sowie die Satzzeichen Punkt (.), Komma (,) und Doppelpunkt (:) wurden im Gegensatz zum Schrägstrich nicht berücksichtigt

Maschinelle Ortsnamenerkennung mit Stanza

Zahlen

- 1.048 Ortsnamen konnten in der Stichprobe identifiziert werden
  - 203 Ortsnamen wurden von Stanza nicht erkannt
  - 114 Ortsnamen davon wurden falsch erkannt

Maschinelle Ortsnamenerkennung mit Stanza

Die wichtigsten Herausforderungen

- Erkennung von Kürzeln und ortsbezogenen Adjektiven
- Erkennung von Ortsangaben, die nicht im Fließtext eingebunden sind
- Erkennung von seltenen Schreibweisen von Ortsnamen
- Verwechslung von Orts-, Personen-, und Objektnamen

Beispiel Jupiter (Meyer, 1788)

- Planet -> Gottheit -> Statue

Zuordnung von Ortskategorien von Ortsnamen in den Brieftexten

- Gebietskörperschaft oder Verwaltungseinheit (932)
  - Staat oder Land (118)
- kleinräumiges Geografikum innerhalb eines Ortes (42)
  - Bauwerk oder Denkmal (39)
- Natürliche geografische Einheit (47)
- Weg, Grenze oder Linie (6)
- Fiktiver Ort (2)
- Extraterrestrikum (16)
- (Innenraum (4))

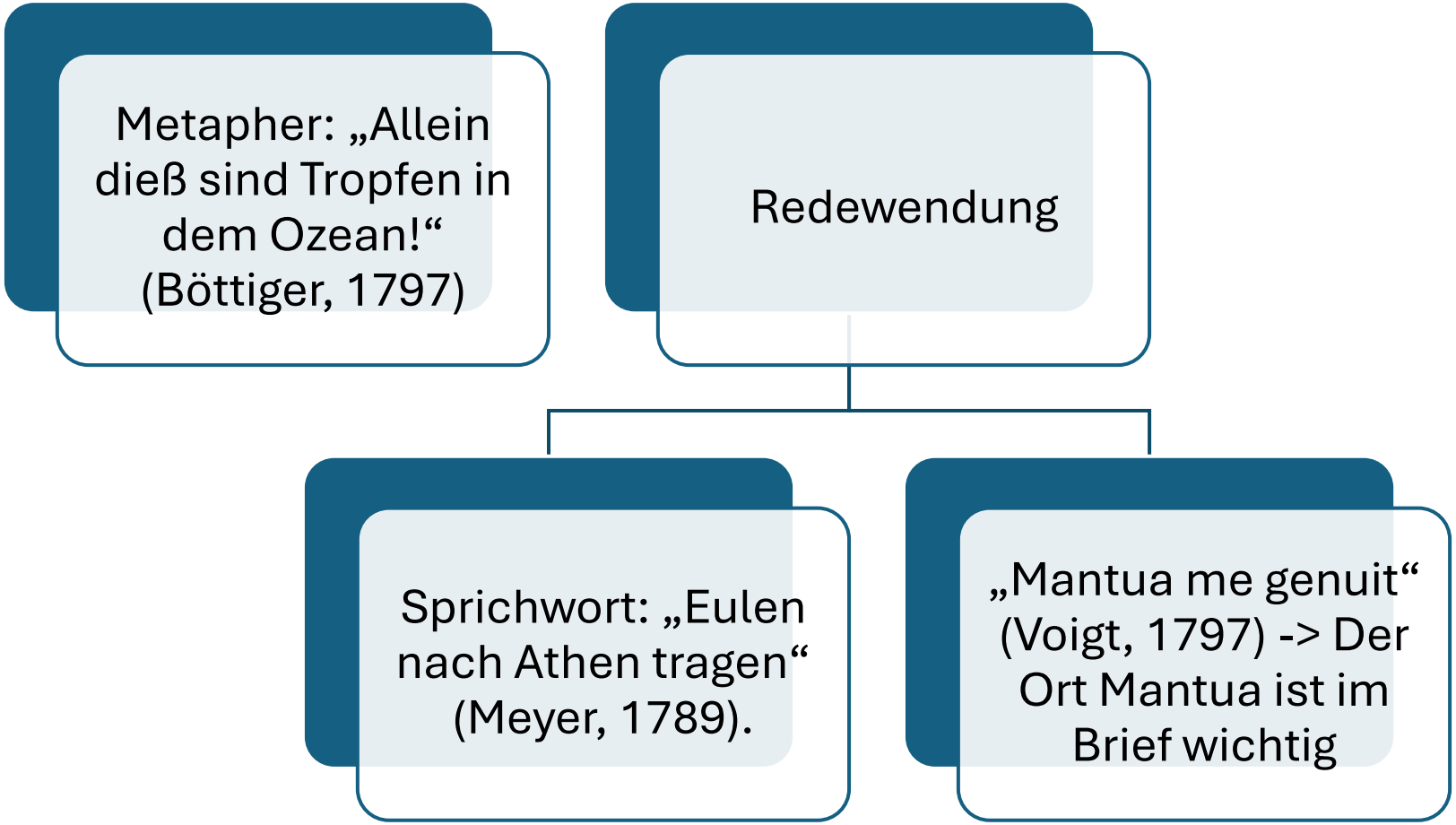
Abstrakte Umschreibungen als Ortsangabe

- Adverbien („hier“, „dort“, „daheim“)
- Adjektive („hiesige“, „vaterländische“)
- Substantive / Gattungsbezeichnungen
  - Unterschiedliche Schreibweisen („Haus“ und „Hauß“)
  - Kürzel („v.H.“)
  - Fremdsprachige Wörter („Cassa“)
  - Zusammengesetzte Wörter („Reithaus“ -> hier ein Ortsname)

Region	**Bauwerk**	**Naturlandschaft**	**Körperschaft**	**Weiteres**
Welt (20)	Haus (50)	Berg (6)	Theater (19)	Ort (22)
Gegend (16)	Schloss (7)	Meer (2)	Bibliothek (5)	Garten (10)
Land (12)	Quartier (6)	Sterne (2)	Gasthof (4)	Weg (2)
Vaterland (10)	Straße (6)	Wüste (1)		Platz (1)
Nation (8)	Hof (5)	Insel (1)		
Stadt (6)	Hütte (5)	Gebirge (1)		
Reich (4)	Kapelle (4)	Feld (1)		
Vaterstadt (4)	Tempel (1)	See (1)		
Dorf (4)	Palast (1)	Wasserfall (1)		
Ausland (3)	Heim (1)	Ozean (1)		
Heimat (1)		Fluss (1)		
	Innenraum			
	Wohnung (10)			
	Zimmer (21)			
	Stube (7)			
	Privatbibliothek (1)			

Tabellarische Darstellung für die in der Stichprobe identifizierten Gattungsbezeichnungen aus den Brieftexten

Metapher und Redewendungen



Ortsangaben als Teil von Ereignissen

- 9 Unterschiedliche Arten von Ereignissen wurden identifiziert und kategorisiert
    1. Brief- und Postsendungen
    2. Aufenthalte und Reisen
    3. Begegnungen
    4. Persönliche Ereignisse
    5. Veranstaltungen
    6. Militärische und politische Ereignisse
    7. Unwetter- und Naturereignisse
    8. Ereignisse in Bezug auf Bauwerke
    9. Fiktive Ereignisse

Unsicherheit und Vagheit im Text beschrieben

Aufenthaltsorte werden vermutet

Aufenthaltsort ist nur ungefähr bekannt

Ortsangaben in Zusammenhang mit anderen Orten, Akteuren, Objekten und Werken

- Ortsangaben können auch für Ereignisse, Körperschaften oder Personen stehen.
- Umgekehrt beinhalten Bezeichnungen für Ereignisse, Körperschaften, oder Personen manchmal eine Ortsangabe.
- Auch können die Entitäten in Relation zueinander stehen

Ortsbeschreibungen

- In 16 Briefen der Stichprobe vom 300 Briefen kommen Ortsbeschreibungen zu Ländern, Städten, Bauwerken und Innenräumen von Gebäuden vor
- Beispiel: Eine Beschreibung von Hamburg

„Der Biblische Ausdruck: Babylon die große Hure! paßt doch auf alle große Städte: ewige Zerstreuung, und ein immerwährendes Lüsteln nur am Schönen und Guten, ohne Samen erweckende Begierde, ohne Sehnsucht und Liebe; aller eigentlichen Lust und Freude kommt man zuvor, wie man dem Hunger u Durste zuvorkommt; das Ganze ist eine Gasterey für lauter verdorbene Magen.“ (Jacobi, 1794)

Diskussion

Herausforderungen bei der Erfassung von Ortsangaben in Briefen:

1. Geeignete Standards oder Richtlinien für die Erschließung von Ortsangaben identifizieren

2. Passende Ortsverzeichnisse, auf die über Identifier verwiesen werden kann, erstellen oder auswählen

3. Ortsnamen maschinell erkennen

4. Abstrakte Umschreibungen von Ortsangaben erschließen, die nur mit Kontextinformation nachvollziehbar sind

5. Ortsangaben als Teil eines Ereignis erschließen

6. Fiktive Orte und Ortsangaben, die Teil eines fiktiven Ereignis sind, erfassen

7. Einen Umgang mit Ortsangaben innerhalb einer Metapher oder einer Redewendung festlegen

8. Relationen von Ortsangaben zu anderen Orten, Personen, Körperschaften, Ereignissen, Werken und Objekten darstellen

9. Bezeichnungen, die mehrere Informationen enthalten, auszeichnen und erschließen. Dazu gehören Personen, Körperschaften und Ereignisse, die gleichzeitig einen Ort meinen und umgekehrt Ortsangaben, die auf Personen, Körperschaften und Ereignisse verweisen.

10. Ortsangaben, die unsicher, vage, fehlerhaft oder nicht identifizierbar sind, kennzeichnen

11. Zuordnung von Koordinaten und Ortskategorien für Orte, insbesondere wenn sich diese im Verlauf der Geschichte ändern

12. Angemessene Visualisierung von Ortsangaben auf einer Karte

13. Identifizierung und Erfassung von Ortsbeschreibungen

Fazit

- Die identifizierten Herausforderungen können Grundlage sein, Ortsangaben zukünftig besser zu erfassen und den Nutzer*innen leichter zugänglich zu machen
- Einen allgemeingültigen Weg zur Lösung der identifizierten Herausforderungen gibt es nicht. Ob und wie diese Herausforderungen in der Praxis bearbeitet werden, muss in Bezug auf Aufwand und Nutzen abgewogen werden

Literatur

- Bosse, A. (2019). II.2 Place. In H. Hotson & T. Wallnig (Hrsg.), Reassembling the Republic of Letters in the Digital Age (S. 79–96). Göttingen University Press.
- Hiltmann, T., Dröge, M., Dresselhaus, N., Grallert, T., Althage, M., Bayer, P., Eckenstaler, S., Mendi, K., Schmitz, J. M., Schneider, P., Sczeponik, W., & Skibba, A. (2025). NER4all or Context is All You Need: Using LLMs for low-effort, high-performance NER on historical texts. A humanities informed approach. https://arxiv.org/abs/2502.04351
- Sahle, P. (2013). Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 2: Befunde, Theorie und Methodik. [Finale Print-Fassung] (Bd. 8). BoD. https://kups.ub.uni-koeln.de/5352/
- Schumacher, M., Flüh, M., & Nantke, J. (2024). Place and Space in Literature: NamedEntity Recognition as a Possibility for Spatial Modelling in Computational Litera-ry Studies. In F. Dammann & D. Kremer (Hrsg.), Geographical Research in the LiteraturDigital Humanities (S. 83–112). Bielefeld University Press. https://doi.org/10.1515/9783839469187-006
- Trzeciak, D. (2023, August). Normdaten und Liebesbriefe: Möglichkeiten und Grenzen der Auszeichnung am Beispiel einer Korrespondenz aus dem Liebesbriefarchiv (S. Bartsch, E. Gius, M. Müller, A. Rapp & T. Weitin, Hrsg.; Bd. 5). TUprints. https://doi.org/ 10.26083/tuprints-00024477
- *Klassik Stiftung Weimar, Sächsische Akademie der Wissenschaften zu Leipzig & Akademie der Wissenschaften und der Literatur Mainz. (2024). PROPYLÄEN. Forschungsplattform zu Goethes Biographica. Goethe- und Schiller-Archiv. Abgerufen am 19. Oktober 2025, **https://goethe-biographica.de/***

Briefe

- Böttiger, K. A. (1797, 20. Juli). Brief an Johann Wolfgang von Goethe: RA 2, Nr. 898 [Version vom 02.09.2024]. In Propyläen Goethes Biographica. KSW / Goetheund Schiller-Archiv; SAW; AdW-Mainz. Abgerufen am 10. Oktober 2025, von http://goethe-biographica.de/id/RA02_ 0898_02474
- Humboldt, W. v. (1796, 25. Juni). Brief an Johann Wolfgang von Goethe: RA 2, Nr. 249 [Version vom 02.09.2024]. In Propyläen Goethes Biographica. KSW / Goetheund Schiller-Archiv; SAW; AdW-Mainz. Abgerufen am 10. Oktober 2025, von http://goethe-biographica.de/id/RA02_ 0249_01823
- Jacobi, F. H. (1794, 16. Dezember). Brief an Johann Wolfgang von Goethe: RA 1, Nr. 1135 [Version vom 02.09.2024]. In Propyläen Goethes Biographica. KSW / Goetheund Schiller-Archiv; SAW; AdW-Mainz. Abgerufen am 10. Oktober 2025, von http://goethe-biographica.de/id/RA01_ 1135_01173
- Meyer, J. H. (1788, 29. Juli). Brief an Johann Wolfgang von Goethe: RA 1, Nr. 264 [Version vom 02.09.2024]. In Propyläen Goethes Biographica. KSW / Goetheund Schiller-Archiv; SAW; AdW-Mainz. http://goethe-biographica.de/id/RA01_ 0264_00296
- Meyer, J. H. (1789, 20. Januar). Brief an Johann Wolfgang von Goethe: RA 1, Nr. 335 [Version vom 02.09.2024]. In Propyläen Goethes Biographica. KSW / Goetheund Schiller-Archiv; SAW; AdW-Mainz. Abgerufen am 10. Oktober 2025, von http://goethe-biographica.de/id/RA01_ 0335_00368
- Voigt, C. G. (1797, Februar). Brief an Johann Wolfgang von Goethe: RA 2, Nr. 607 [Version vom 02.09.2024]. In Propyläen Goethes Biographica. KSW / Goetheund Schiller-Archiv; SAW; AdW-Mainz. Abgerufen am 10. Oktober 2025, von http://goethe-biographica.de/id/RA02_ 0607_02182

Vielen Dank für Ihre Aufmerksamkeit!