




# Digitale Harmonie aus historischer Dissonanz

---

Extraktion, Ordnung und Analyse  
unstrukturierter Archivdaten  
des Männerchor Murg

Sven Burkhardt

 0009-0001-4954-4426

 17-056-912

 15.08.2025



University  
of Basel



Digital  
Humanities  
Lab

University of Basel  
Digital Humanities Lab  
Switzerland



# Abstract

Diese Arbeit befasst sich mit dem Archiv des Männerchor Murg in den Jahren des Zweiten Weltkrieges. Hierfür wird eine automatisierte Pipeline auf Basis von LLMs und Pattern-matching vorgestellt, mit deren Hilfe Named Entities extrahiert und weiterverarbeitet werden. Ziel ist es, dieses Archiv digital zugänglich zu machen, die beteiligten Personen sowie deren Netzwerke und dessen geographische Ausdehnung sichtbar zu machen.

# Inhaltsverzeichnis

<b>Einleitung</b>	<b>1</b>
Ziel und Relevanz der Arbeit . . . . .	1
Forschungsstand und Forschungslücke . . . . .	1
Formulierung der Forschungsfrage . . . . .	1
Aufbau der Arbeit . . . . .	1
Geografischer und historischer Kontext . . . . .	1
<b>Quellen</b>	<b>2</b>
Quellentradierung . . . . .	2
Quellenbeschrieb . . . . .	2
Zeitraum . . . . .	2
Dokumententyp . . . . .	2
Inhalt??? . . . . .	2
<b>Korpus</b>	<b>2</b>
Digitalisierung . . . . .	3
Sichtung & Kategorisierung in Akten . . . . .	3
Transkription . . . . .	3
Test mit Tesseract . . . . .	3
Test mit LLM . . . . .	3
Transkribus . . . . .	4
Tagging . . . . .	4
Tagging mit Transkribus . . . . .	4
Tagging mit LLM . . . . .	4
Export . . . . .	4
<b>Methodenkritik</b>	<b>5</b>
Genutze Tools . . . . .	5
Msty . . . . .	5
Alphabet – Gemini . . . . .	5
Anthopic – Claude . . . . .	5
OpenAI – ChatGPT . . . . .	5
Transkribus . . . . .	5
Nodegoat . . . . .	5
Netzwerkanalyse als Methode . . . . .	5
Theoretischer Hintergrund der Netzwerkanalyse . . . . .	5
Ziele der Netzwerkanalyse im Kontext der Quellen . . . . .	5
Technische Umsetzung (Tools, Datenbankstruktur) . . . . .	5
<b>Pipeline</b>	<b>5</b>
Aufbau XML to JSON Pipeline . . . . .	5
Übersichtsgrafik der Pipeline . . . . .	5
Module im Detail . . . . .	5
document_schemas.py . . . . .	5
__init__.py . . . . .	5
Person-Matcher . . . . .	7
place_matcher.py . . . . .	8
organization_matcher.py . . . . .	8
letter_metadata_matcher.py . . . . .	8

type_matcher.py . . . . .	8
event_matcher.py . . . . .	8
date_matcher.py . . . . .	8
Assigned_Roles_Module.py . . . . .	8
unmatched_logger.py . . . . .	8
KEINE AHNUNG WAS DIE HIER MACHEN . . . . .	8
validation_module.py . . . . .	8
validation_module.py . . . . .	8
test_role_schema.py . . . . .	8
llm_enricher.py . . . . .	8
enrich_pipeline.py . . . . .	8
<b>Analyse &amp; Diskussion der Ergebnisse</b>	<b>8</b>
Visualisierung auf der VM . . . . .	8
<b>Fazit und Ausblick</b>	<b>8</b>
Zusammenfassung der zentralen Erkenntnisse . . . . .	8
Methodische Herausforderungen und Lösungen . . . . .	8
Ausblick auf zukünftige Forschung und mögliche Erweiterungen der Datenbank .	8
<b>ALTER SCHEISS</b>	<b>9</b>
Forschungsstand zu den Quellen . . . . .	9
Beschreibung des Archivbestands . . . . .	10
<b>Methodischer Zugang</b>	<b>11</b>
Digitale Erfassung und Strukturierung der Quellen . . . . .	11
Gliederung in Akten . . . . .	11
Digitalisierung und Transkription . . . . .	11
Tagging in Transkribus . . . . .	11
Strukturelle Tags . . . . .	12
Inhaltliche Tags . . . . .	13
Digitalisierungsprozess und Herausforderungen . . . . .	15
Diagram english . . . . .	21
Gründe für den Wechsel zu Nodegoat . . . . .	23
Nodegoat Modellierung . . . . .	23
Netzwerkanalyse als Methode . . . . .	23
Theoretischer Hintergrund der Netzwerkanalyse . . . . .	23
Ziele der Netzwerkanalyse im Kontext der Quellen . . . . .	23
Technische Umsetzung (Tools, Datenbankstruktur) . . . . .	23
Normalisierung der Dateien — von PDF zu JPEG . . . . .	23
<b>Aufbau der Datenbank</b>	<b>25</b>
Konzeption der Datenmodellierung . . . . .	25
Eigene Ontologie im Vergleich zu bestehenden Standards . . . . .	25
Verknüpfung von Personen, Orten und Ereignissen . . . . .	25
Implementierung der Datenbank . . . . .	25
Datenbankdesign . . . . .	25
Herausforderungen bei der Datenaufnahme . . . . .	25
Verknüpfung mit externen Quellen (z.B. Wikidata) . . . . .	25
<b>Analyse der Netzwerke</b>	<b>26</b>

Soziale Netzwerke des Vereinslebens . . . . .	26
Verbindungen zwischen Mitgliedern . . . . .	26
Kooperationen mit anderen Vereinen . . . . .	26
Politische Netzwerke und deren Veränderungen . . . . .	26
Einfluss der NS-Diktatur auf die Netzwerke . . . . .	26
Feldpostkarten als Quelle für militärische Netzwerke . . . . .	26
Geografische Ausdehnung der Netzwerke . . . . .	26
Einsatzorte der Chormitglieder während des Krieges . . . . .	26
Lokale und überregionale Verbindungen . . . . .	26
<b>Diskussion der Ergebnisse</b>	<b>27</b>
Sichtbarmachung der Netzwerke durch Nodegoat und Netzwerkanalyse . . . . .	27
Gibt es Veränderungen der Netzwerke im historischen Kontext? . . . . .	27
<b>Bibliographie</b>	<b>28</b>
<b>References</b>	<b>28</b>

# Einleitung

## Ziel und Relevanz der Arbeit

## Forschungsstand und Forschungslücke

## Formulierung der Forschungsfrage

## Aufbau der Arbeit

## Geografischer und historischer Kontext

Die vorliegende Arbeit beschäftigt sich mit Unterlagen aus dem Archiv des "Männerchor Murg" und dessen Nachfolger, den "New Gospelsingers Murg". Murg ist eine Gemeinde am Hochrhein, rund 30 km Luftlinie von Basel entfernt. Der Ort liegt am gleichnamigen Fluss Murg, der in den Rhein mündet. Beide Gewässer bildeten über Jahrhunderte hinweg den wirtschaftlichen Motor der Region: Die Wasserkraft der Murg begünstigte früh die Ansiedlung von Mühlen, Hammerwerken und Schmieden entlang des Bachlaufs. Parallel bot der Rhein mit seiner Drahtseil-Fähre eine wichtige Verkehrs- und Handelsverbindung, und bis zum Ersten Weltkrieg privat betrieben wurde.

Mit dem Ausbau der Landstraße, der heutigen Bundesstraße 34, sowie dem Anschluss an die Bahnstrecke Basel–Konstanz wandelte sich Murg im 19. Jahrhundert von einer landwirtschaftlich geprägten Siedlung zu einer Gewerbe-, Handels- und Industriegemeinde. Während der Industrialisierung Mitte des 19. Jahrhunderts wurde die Wasserkraft zu einem entscheidenden Faktor der Region. Die Ansiedlung der Schweizer Textilfirma *Hüssy & Künzli AG* im Jahr 1853<sup>1</sup> trug wesentlich zum Wachstum der Gemeinde bei. Zahlreiche Arbeitskräfte — auch aus der benachbarten Schweiz — machten die Gemeinde zu einem wichtigen Standort der regionalen Textilindustrie. Es waren eben diese schweizer Textilarbeiter, die 1861 den *Männerchor Murg* gründeten, damals unter dem Namen *Schweizer Männerchor Murg*.

---

1. Vgl. [Gemeinde Murg](#),

# Quellen

## Quellentradierung

In den Lagerräumen der New Gospel Singers Murg, dem Nachfolgeverein des Männerchors Murg, wird im Jahr 2018 mehrere je ca. 800 Seiten umfassende Ordner mit historischen Unterlagen gefunden. Für diese Arbeit wird ein Ordner mit der Aufschrift *“Männerchor Akten 1925-1944”* gewählt, da er neben dem Ordner *“Männerchor Akten 1946-1950”* den größten Zeitraum abdeckt. Darüberhinaus bietet er das Potential, aufschlussreiche Einblicke in das Vereinsleben in der Zeit vor und während des Nationalsozialismus, insbesondere des Zweiten Weltkrieges, zu geben.

Der Ordner umfasst insgesamt 780 Seiten und deren Inhalt kann als “Protokoll”, “Brief”, “Postkarte”, “Rechnung”, “Regierungsdokument”, “Noten”, “Zeitungsartikel”, “Liste”, “Notizzettel” oder “Offerte” kategorisiert werden.

## Quellenbeschreibung

### Zeitraum

blabla

### Dokumententyp

blabla

### Inhalt???

## Korpus

Aus dem Bestand des Ordners *“Männerchor Akten 1925-1944”* werden für diese Arbeit ausschließlich Akten verwendet, die während des Zweiten Weltkriegs verfasst wurden. Der Analysezeitraum erstreckt sich dementsprechend zwischen dem 01. September 1939 und dem 8. Mai 1945<sup>2</sup>, dem Tag der bedingungslosen Kapitulation Deutschlands.

Diese Begrenzung des zeitlichen Rahmens ist notwendig, um die Funktionalität der weiter unten beschriebenen Pipeline darstellen zu können. Hieraus ergibt sich in der Folge eine Limitation der Anzahl potentieller Akteurinnen und Akteure, Orte und Organisa-

---

2. [\[\[vgl.\]\]](#)[\[Finde hier eine Referenz\]](#)keylist

tionen. Notwendig ist das besonders mit Hinblick auf die Erstellung einer verlässlichen Groundtruth mit angereicherten historischen Daten durch Archivrecherchen. Sie helfen, das Potential solch einer digital unterstützten Arbeitsweise zu unterstreichen.

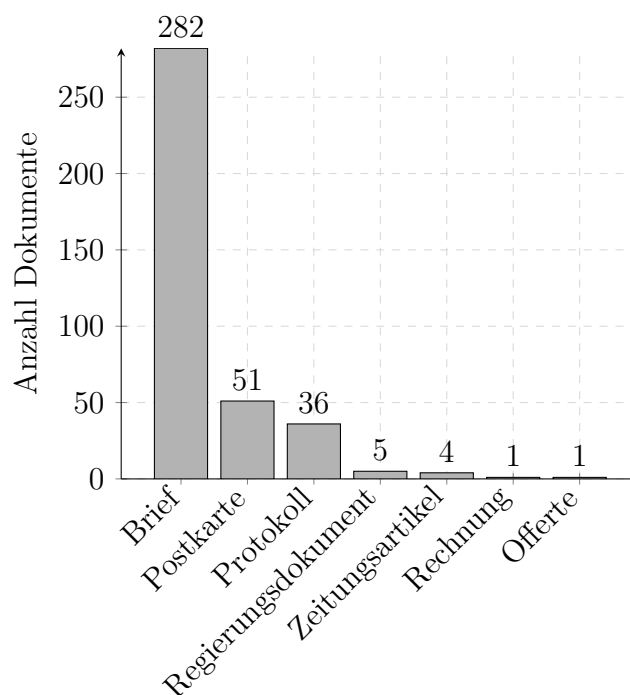


Abbildung 1: Verteilung der Dokumententypen im untersuchten Bestand (150 Akten - 381 Seiten).

## Digitalisierung

blabla

## Sichtung & Kategorisierung in Akten

blabla

## Transkription

blabla

## Test mit Tesseract

blabla

## Test mit LLM

blabla



## **Transkribus**

blabla

## **Tagging**

blabla

## **Tagging mit Transkribus**

blabla

## **Tagging mit LLM**

blabla

## **Export**

blabla

# Methodenkritik

## Genutze Tools

Msty

Alphabet – Gemini

Anthropic – Claude

OpenAI – ChatGPT

Transkribus

Nodegoat

## Netzwerkanalyse als Methode

Theoretischer Hintergrund der Netzwerkanalyse

Ziele der Netzwerkanalyse im Kontext der Quellen

Technische Umsetzung (Tools, Datenbankstruktur)

## Pipeline

### Aufbau XML to JSON Pipeline

Übersichtsgrafik der Pipeline

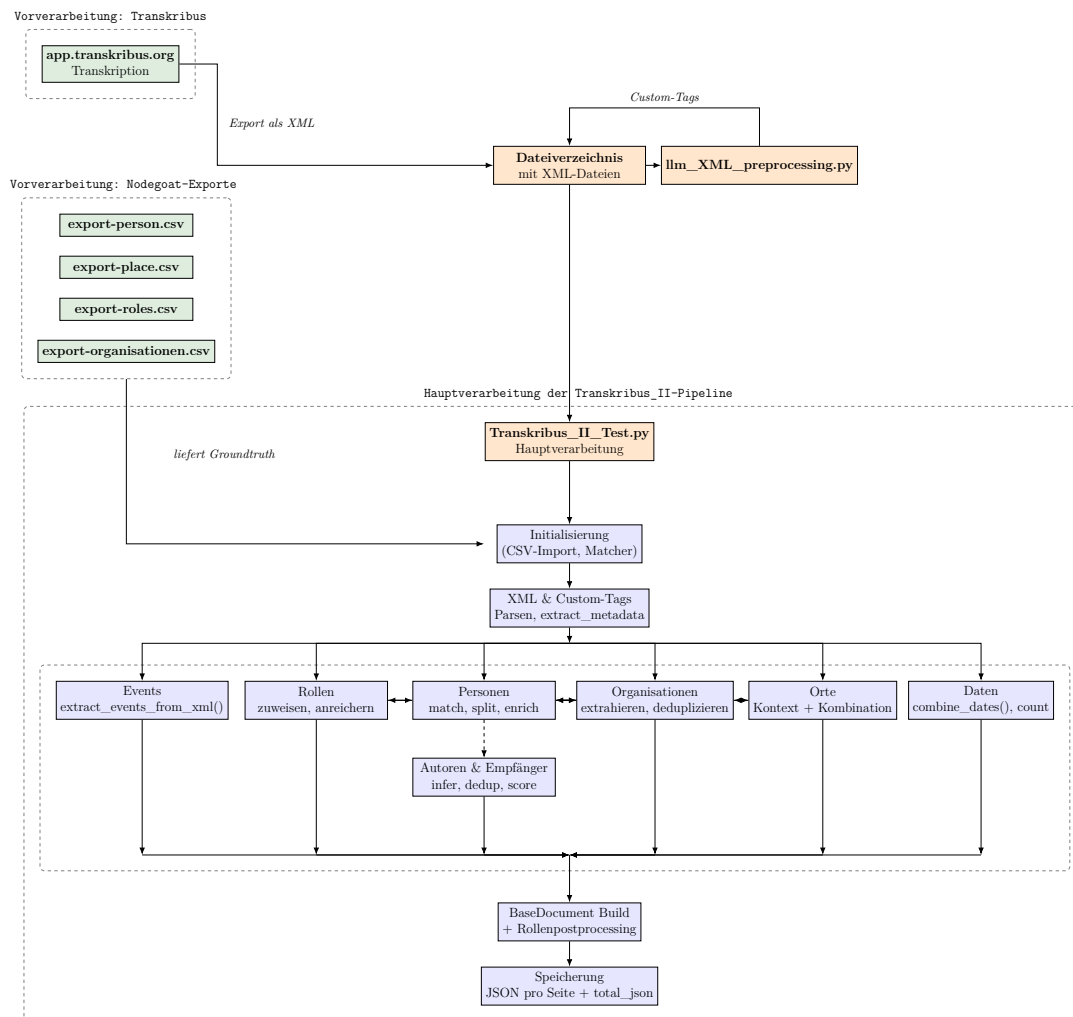


Abbildung 2: Übersicht der gesamten XML-to-JSON-Pipeline

## Person-Matcher

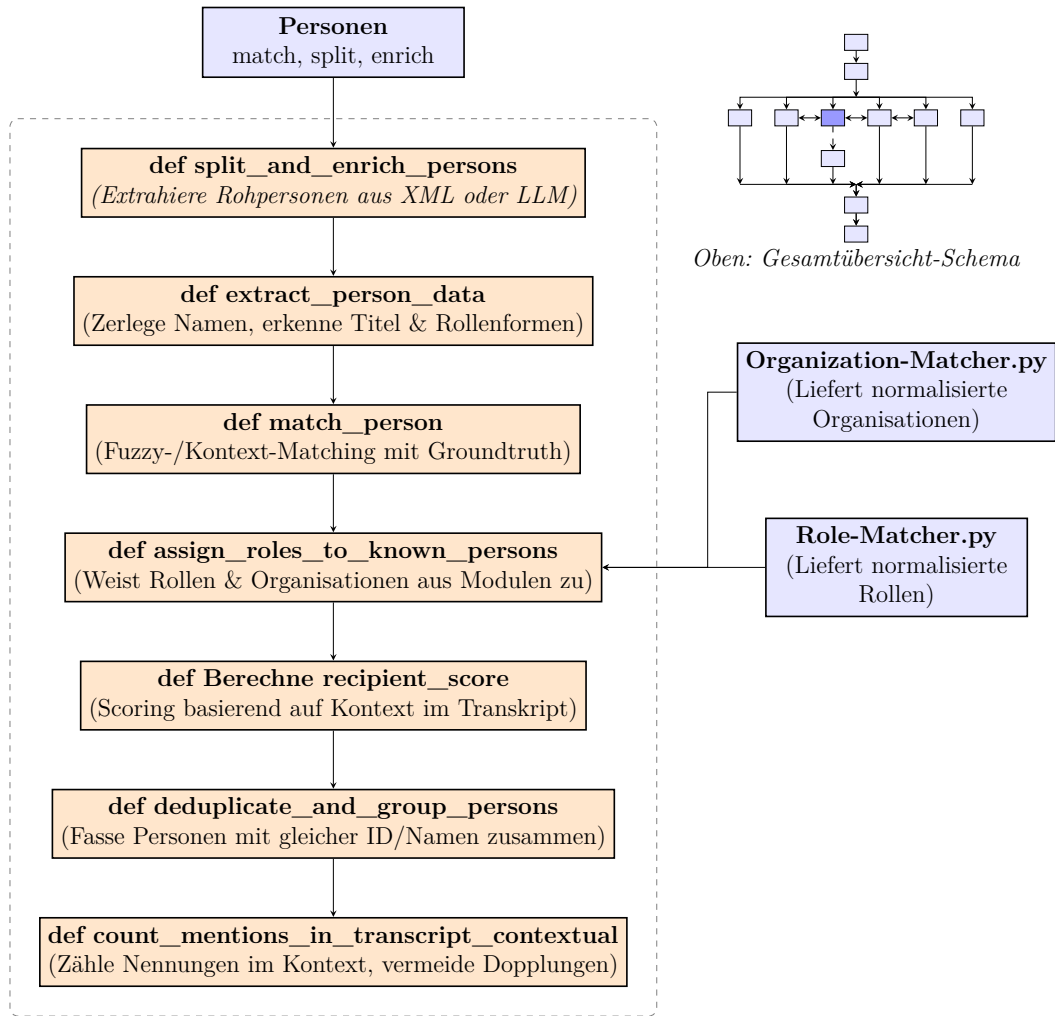


Abbildung 3: Detailliertes Prozessdiagramm: Personen-Matching

place\_\_matcher.py

organization\_\_matcher.py

letter\_\_metadata\_\_matcher.py

type\_\_matcher.py

event\_\_matcher.py

date\_\_matcher.py

Assigned\_Roles\_Module.py

unmatched\_\_logger.py

**KEINE AHNUNG WAS DIE HIER MACHEN**

validation\_\_module.py

validation\_\_module.py

test\_\_role\_\_schema.py

llm\_\_enricher.py

enrich\_\_pipeline.py

**Analyse & Diskussion der Ergebnisse**

Visualisierung auf der VM

**Fazit und Ausblick**

Zusammenfassung der zentralen Erkenntnisse

Methodische Herausforderungen und Lösungen

Ausblick auf zukünftige Forschung und mögliche Erweiterungen  
der Datenbank

## ALTER SCHEISS

In Transkribus-Seminaren am Departement Geschichte der Universität Basel wird aus *“Männerchor Akten 1925-1944”* bereits 2018 und 2022 ein erster Korpus aus 137 Akten<sup>3</sup>. Es entsteht eine Liste, die die Seiten mit deren Lage im Ordner, einem Kurztitel und einem Entstehungsdatum versieht. Als Akte werden im Folgenden Schriftstücke bezeichnet, die entweder durch die Fundsituation, oder ihren Inhalt eindeutig als zusammengehörig betrachtet werden können. So liegt Akte\_001 beispielsweise in einer separaten Mappe und umfasst 96 Seiten, während andere Akten nur aus einer einzelnen Seite bestehen können. Während der Fokus 2028 auf den augenscheinlich häufig auftretenden Personennamen "Carl Burger" und "Fritz Jung" liegt, wird 2022 im Rahmen eines zweiten Seminars spezifischer die Feldpost untersucht. Zu diesem Zeitpunkt erfolgt die Transkription mit einem generischen Modell, das nicht auf die unterschiedlichen Handschriften trainiert ist.

## Forschungsstand zu den Quellen

Die vorliegende Arbeit stützt sich auf diese Vorarbeit, und die darin gesammelten Daten. Beispielsweise werden die Feldpostbriefe um weitere Daten ergänzt. Kernfragen hierfür sind: *Welche Einheiten verbergen sich hinter den Feldpostnummern? Wo waren die Einheiten, als der Brief geschrieben werden?*

Hierzu werden Nachschlagetabellen in Fachliteratur<sup>4</sup>, die Bestände des *Bundesarchives – Militärarchiv Freiburg*<sup>5</sup>, des *Suchdienstes des Deutschen Roten Kreuzes (DRK)*<sup>6</sup>, sowie Citizen Scientist Projekte<sup>7</sup> herangezogen und letztere teils durch eigene Recherche ergänzt.

Für diese Arbeit wird die Kategorisierung von 2018 übernommen und auf den Seiten im Ordner erweitert.<sup>8</sup>

---

3. Weiterführend vgl. (Burkhardt 2022)

4. vgl.: (Tessin 1977), (Hartmann 2010), (Rass und Rohrkamp 2009)

5. Hollmann 2025.

6. „DRK Suchdienst | Suche per Feldpostnummer“ 2025.

7. vgl. Wikidata (Beispiel: („78th Sturm-Division (Wehrmacht)“ 2024)),

(„Lexikon der Wehrmacht“, o. Dat.),

(„Forum Geschichte der Wehrmacht“, o. Dat.)

8. Akten\_Gesamtübersicht.csv im Anhang

## Beschreibung des Archivbestands

## Methodischer Zugang

### Digitale Erfassung und Strukturierung der Quellen

#### Gliederung in Akten

Die analogen Akten müssen zuerst für die Digitalisierung vorbereitet werden. Sie werden aus den Ordnern genommen und vorsichtig von Heftklammern, Gummibändern und Büroklammern befreit. Dies dient der Konservierung des Papiers – gerade an Stellen, an denen sich vorher Büroklammern befunden haben, frisst sich Rost in das Papier und beschädigt es stark. Auch sonstiger Säurefraß durch nicht-säurefreies Papier, das sich im Ordner befand, zeigt sich an einigen Stellen.

Um schnell und dennoch in guter Auflösung zu digitalisieren, wird die „Dateien“-App<sup>9</sup> von Apple benutzt, da sie gleichzeitig einen grossen Cloud-Speicher und eine OCR-Erkennung bietet. Die Intention dahinter sind schnell durchsuch- und auffindbare Texte. Um die Geschwindigkeit der Digitalisierung zu erhöhen, und eine vergleichbare Qualität zu erhalten, wird ein Ipad mit einem Stativ verwendet, das im 90°Winkel über den Seiten positioniert ist. Die Dateien werden entsprechend der bereits erwähnten Akten\_Gesamtübersicht benannt. Sind mehrere Blätter zusammengeheftet, so ergeben sie eine Akte. Sind sie einzeln, werden sie ebenfalls als einzelne Akte geführt. Die Archivierung findet sowohl analog wie digital auf Seiten-Ebene statt.

#### Digitalisierung und Transkription

#### Tagging in Transkribus

Transkribus und seine Modelle unterstützen nicht nur beim Transkribieren der Texte, sondern erlauben auch das Taggen von *Named Entities*. Für die vorliegende Arbeit sind dabei besonders Personen, Orte, Organisationen und Daten relevant. Um hierfür ein stringentes Verfahren zu entwickeln, wurden die Tags wie folgt definiert:

---

9. vgl. [Apple-Finder](#)



## Strukturelle Tags

### abbrev

Mit dem Tag **abbrev** werden alle Abkürzungen getaggt, die für eine eindeutige Entität stehen.

☞ **Beispiel 1:** Dr., Prof., St., Hr., Frl., Dipl.-Ing., etc.

☞ **Beispiel 2:** Organisationskürzel, wenn sie eindeutig sind:  
„<abbrev>V.D.A.</abbrev>“.

☞ **Beispiel 3:** Falls eine ausgeschriebene Variante im selben Dokument vorhanden ist, bleibt die Abkürzung getaggt:

```
<person><abbrev>Dr.</abbrev>Weiß</person>
```

### unclear

Mit dem Tag **unclear** werden unleserliche oder schwer entzifferbare Textstellen markiert.

☞ **Beispiel 1:** Unklare Zeichen oder fehlende Buchstaben:

```
„Er wohnte in<unclear>[...]<unclear>“.
```

☞ **Beispiel 2:** Teilweise lesbare Wörter:

```
"<place>Frei<unclear>[...]<unclear><place>“.
```

### sic

Mit dem Tag **sic** werden Wörter markiert, die im Originaltext in einer falschen oder ungewöhnlichen Schreibweise geschrieben wurden.

☞ **Beispiel 1:** Veraltete oder falsche Schreibweisen:

```
„<sic>daß</sic>“ für dass.
```

☞ **Beispiel 2:** Offensichtliche Tippfehler, wenn sie im Originaltext so vorkommen:

```
„Wir haben <sic>einen</sic> große Freude.“
```

☞ **Beispiel 3:** Falls eine Korrektur notwendig ist, kann sie als Kommentar ergänzt werden.

## Inhaltliche Tags

### person

Mit dem Tag **person** sollen alle Strings getaggt, die eine direkte Zuordnung einer Person ermöglichen.

☞ **Beispiel 1:** Vereinsführer, Alfons, Zimmermann, Alfons Zimmermann, Z. A. Zimmermann, Herr Zimmermann, Herr Alfons Zimmermann, etc.

☞ **Beispiel 2:** Funktionen wie Oberlehrer, Chorleiter, etc., wenn Ort, Name oder Organisation bekannt.

Eine Person kann sowohl mit ihrem Namen als auch ihrer Funktion (wie Dirigent) getaggt werden. Aus der Korrespondenz ist in der Regel eine zugehörige Organisation ersichtlich, mit deren Verknüpfung eine namentlich nicht genannte Person identifiziert werden könnte.

### signature

Mit dem Tag **signature** werden alle Strings getaggt, die eine handschriftliche Unterschrift darstellen. Der Tag **signature** ist nahezu deckungsgleich mit dem Tag **person**. Er dient zur **graduellen Unterscheidung**, ob ein Name im Fließtext als gesichert leserlich oder handschriftlich als Signatur vorliegt.

☞ **Beispiel 1:** Eindeutig lesbare Signaturen werden direkt getaggt:

```
<signature>A. Zimmermann</signature>.
```

☞ **Beispiel 2:** Teilweise unleserliche Signaturen werden mit dem Tag **unclear** innerhalb von **signature** markiert:

```
<signature>R. We<unclear>[...]</unclear></signature>.
```

☞ **Beispiel 3:** Wenn nur ein Teil des Namens lesbar ist, aber eine Identifikation unsicher bleibt, sollte die Unterschrift vollständig im Tag **unclear** innerhalb von **signature** stehen:

```
<signature><unclear>Unleserlich</unclear></signature>.
```

☞ **Beispiel 4:** Wenn eine Signatur einer bekannten Person zugeordnet werden

kann, aber nicht vollständig lesbar ist, bleibt die Signatur erhalten und wird **ohne** den Tag **person** zu verwenden:

```
<signature>A. Zimm<unclear>[...]</unclear></signature>.
```

☞ **Beispiel 5:** Wenn eine Unterschrift vollständig transkribiert wurde und die Person bekannt ist, wird sie nur mit **signature** getaggt, **ohne** den Tag **person** zu verwenden: 

```
<signature>Alfons Zimmermann</signature>.
```

## organization

Mit dem Tag **organization** werden alle Strings getaggt, die eine direkte Zuordnung einer Organisation ermöglichen.

☞ **Beispiel 1:** Männerchor Murg, Verein Deutscher Arbeiter (V.D.A.), Murgtalschule, etc.

☞ **Beispiel 2:** Abkürzungen, wenn sie eine Organisation eindeutig bezeichnen, z.B. V.D.A., NSDAP, STAGMA, etc.

## place

Mit dem Tag **place** werden alle Strings getaggt, die sich auf einen geografischen Ort beziehen.

☞ **Beispiel 1:** Murg (Baden), Freiburg, Berlin, Murgtal, Schwarzwald, etc.

☞ **Beispiel 2:** Orte mit näherer Bestimmung, z.B. „bei Berlin“, „im Murgtal“ werden getaggt:

```
<place>im Murgtal</place>.
```

## date

Mit dem Tag **date** werden alle expliziten und implizierten Datumsangaben markiert.

☞ **Beispiel 1:** 29.05.1936

☞ **Beispiel 2:** 29. Mai 1936

☞ **Beispiel 3:** den 29. d. Mts.:

```
<date when="29.05.1936 ">den 2.</date> <abbrev>d. Mts.</abbrev>
```

## event

Mit dem Tag **event** werden expliziten und implizierten Ereignisse markiert. Diese Ereignisse haben einen zeitlichen oder räumlichen Bezug, und können benannt werden. Dazu zählen: ☞ **Beispiel 1:** "Jubiläumskonzert"

☞ **Beispiel 2** "Gründung des Vereins"

☞ **Beispiel 2** "Kriegsausbruch" oder "Kriegsende"

Konzepte, die nicht klar in den Texten benannt werden, wie beispielsweise die Suche nach einem Dirigenten, können nicht immer Ereignis getaggt werden. Sie sollen später aber in der Datenbank implementiert werden.

## Digitalisierungsprozess und Herausforderungen

Hier gehört dringend dazu, dass die Quellen über einen längeren Zeitraum digitalisiert wurden. Das bedeutet, dass sich die Kameras geändert haben. Verwendet wurden primär ein iPad Pro 2nd Generation (2017) und ein iPad Air 4th Generation (2022). Die Verwendete Software ist die Scan-Funktion von Apple iCloud. Die Auswahl der Software war aus rein ökonomischen Gründen. Da das Digitalisierungsprojekt bereits 2018 begonnen wurde, fehlten weitestgehend Grundlagenkenntnisse, die im Digital Humanities Studium vermittelt wurden. Berücksichtigt wurden jedoch einige Richtlinien, wie sie in den Archiv-Kursen des Bachelor-Geschichtsstudiums vermittelt wurden (gleichbleibende Beleuchtung, Hintergrund). Die Scanqualität ist daher oft nicht optimal, was zu Problemen bei der OCR Erkennung mit OCR Software (Apple OCR, Adobe, etc.) führte. Aus diesem Grund wurden 75 Akten zunächst mit dem Model "The German Giant I" mit einer CER von 8,30% transkribiert. Mit insgesamt 4 Iterationen wurde eine Groundtruth für ein eigenes Modell erstellt, und gleichzeitig Personen, Orte, Daten und Organisationen getaggt. Hierzu wurde auch manuell OpenAIs CHatGPT 4o Modell verwendet, das für die Rechtschreibprüfung verwendet wurde. Tauchte ein Rechtschreibfehler im Text auf, wurde dieser manuell überprüft. War der Fehler bereits im Ursprungstext, so wurde der Tag "sic" verwendet, und eine Korrektur beigefügt.

Die so erstellten 70 Akten ergaben 158 Seiten zu insgesamt 22.155 Wörtern Groundtruth,

womit dann ein eigenes Transkribus Modell<sup>10</sup> (ModelID: 287793) erstellt wurde. Es erreichte eine Accuracy (CER) von 6,58%. Später wurden die verbleibenden 80 Akten nur noch mit diesem Modell transkribiert.

ChatGPT produziert daraus:

Durch CHatGPT verliert der Text zwar seine ursprüngliche Formatierung und Zeilenumbrüche, aber wird nun nahezu fehlerfrei lesbar. Nur das "Venstadler Liedchen" ist eigentlich eines aus "Neustadt". Eine anschließende menschliche Korrektur ermöglicht also den Abgleich mit dem nun lesbaren Text, und die Korrektur der Transkription.

Korrigiert und getagt lautet der Brief nun:

München, 28.V.1941

Lieber Otto!

Nur wer die Sehnsucht kennt weiß was ich leide

Ich wandle traurig her in schwarzer Seide.

Die Sehnsucht brennt, du bist so fern.

Ach lieber Otto, wie hab ich dich gern.

Ich schnitt es gern in alle Rinden.

Ach Otto, wann u. wo kann ich dich finden?

Deine dich nie vergessende

Lina Fingerdick

An

Herrn Otto Bollinger

z.Hd. Herrn Alfons Zimmermann

Vereinsführer des Männerchor

Murg

Laufenburg (Baden)

Rhina

---

10. burkhardt\_transkribus\_nodate.

Münch, 15. Aug. 41.

Mein lieber Hans!

Dieser lange Brief ist mir aus Män-  
chen wieder einmal ein Liebesgut zu fließen.  
und kann mir die gottliche Gabe zufließen.

Es ist mir so wunderbar, dass aus dem  
Mannchen die Stadt um den Titel des Liebes-  
gutes zu erhalten, so für einen Abschied aus der  
Stadt „aus der Stadt“ über oft in der  
Stadt beifügt, kein Abschied. Vielleicht  
gelingt es dir diesen Titel zu erhalten.

Dieses ist die Stadt in der Stadt  
„das selbe Lied“ von dem. Nicht

Es wurde 1928 um 10. Reichs. Längst. Post  
um Längst. Längst. in der Stadt  
und wurde in der Stadt.

Es ist so, dass die Stadt zu finden.

Als Hans: „das selbe Lied“ von dem,  
als Hans: „das selbe Lied“ von dem,  
Lied ist unmöglich, dass es die Stadt.

Mit sehr. Hans

Ein

Carl.

Abbildung 4: Beispiel für handschriftlichen Text in Akte\_076 erkannt mit Transkribus

Murg. 15. Aug 41

Mein lieber Alfons!  
Sehen lunge Leitt es mich dem Männer-  
chor wieder einmal ein Liedchen zu stehen.  
und kam mir die gestege Gelegenheit gussend.  
Männechor Venstad um den Title das Liedchen  
zu erhalten, wo sie zum Abschied am Aute  
sängen „auf Wiederschen Owohl ich Frei!  
märke beifügte, keine Aentwarb. Vielleicht  
gelingt es Dir diesen litel zu erhalten.  
Weiterhin sänge ich fal Lied nur  
"Bas alte Lied von being. Rerohl  
Es wurde 1928 am 10. Dachub. Sängerb. Frst  
von Begrüßungsabend in Dien gesungen.  
und erntete überaus großen Reifall.  
Es ich schwer das Richtige zu finden.  
Aler Alfon, werst das Vemsladler Liedchen.  
alsdann das Biener Lidchen und wenn  
Leides unmöglich, dann freu Nall.  
Mit herzl. Grüße  
Dein  
Carl

Abbildung 5: Transkription von [Abbildung 4](#)



Murg, 15. Aug. 41

Mein lieber Alfons!

Schon lange treibt es mich, dem Männerchor wieder einmal ein Liedchen zu stiften, und kam mir die günstige Gelegenheit gelegen.

Ich schrieb vergangenes Jahr an den Männerchor Venstad, um den Titel des Liedchens zu erhalten, das sie zum Abschied am Auto sangen: „Auf Wiedersehen, o wohl ich frei!“

Ich fügte eine Frankierung bei, erhielt jedoch keine Antwort. Vielleicht gelingt es Dir, diesen Titel zu erhalten.

Weiterhin sang ich das Lied nur „Das alte Lied von Wien“. Obwohl es am 10. Dezember 1928 beim Sängerbund-Fest von Begrüßungsabend in Wien gesungen wurde und überaus großen Beifall erntete, ist es schwer, das Richtige zu finden.

Aber Alfons, zuerst das Venstadler Liedchen, dann das Wiener Liedchen und wenn beides unmöglich, dann Fröhlichsein.

Mit herzlichen Grüßen

Dein

Carl

Abbildung 6: Transkription durch ChatGPT von [Abbildung 5](#)



Murg. 15. Aug 41

Mein lieber Alfons!

Seit langem treibt es mich dem Männer-chor wieder einmal ein Liedchen zu stiften. und kam mir die günstige Gelegenheit passend.

Ich schrieb vergangenes Jahr an den Männechor Vorstand um den Titel das Liedchen zu erhalten, wo sie zum Abschied am Auto sangen "auf Wiederschen" Obwohl ich Frank-marke beifügte, keine Antwort. Vielleicht gelingt es Dir diesen Titel zu erhalten.

Weiterhin sänge ich das Lied nur

"Das alte Lied" von Komp. Kirchl

Es wurde 1928 am 10. Deutsch. Sängerb. Fest am Begrüßungsabend in Wien gesungen.

und erntete überaus großen Beifall.

Es ist schwer das Richtige zu finden.

Also Alfons! zuerst das Neustadter Liedchen.

alsdann das Wiener Liedchen und wenn

Beides unmöglich, dann freie Wahl.

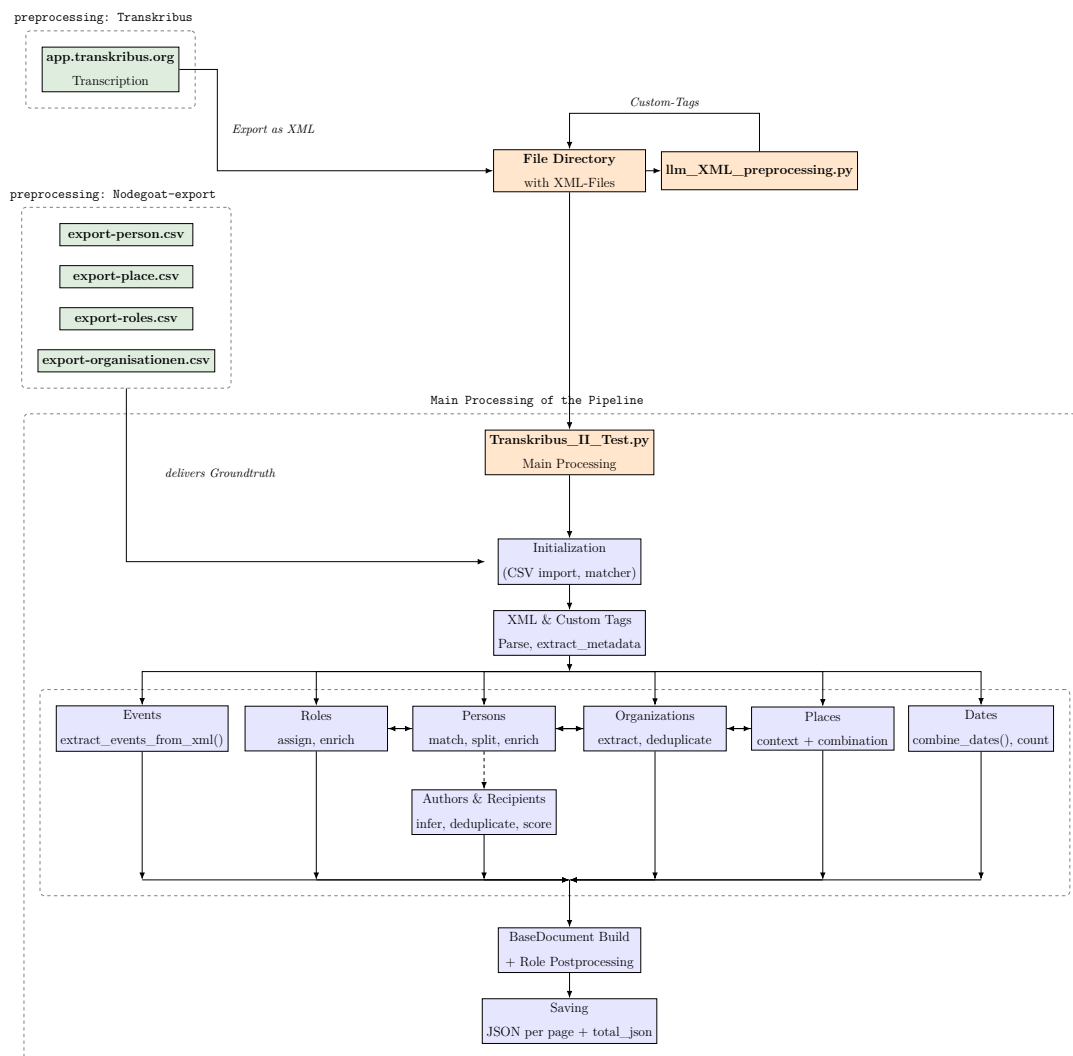
Mit herzlichen Grüßen

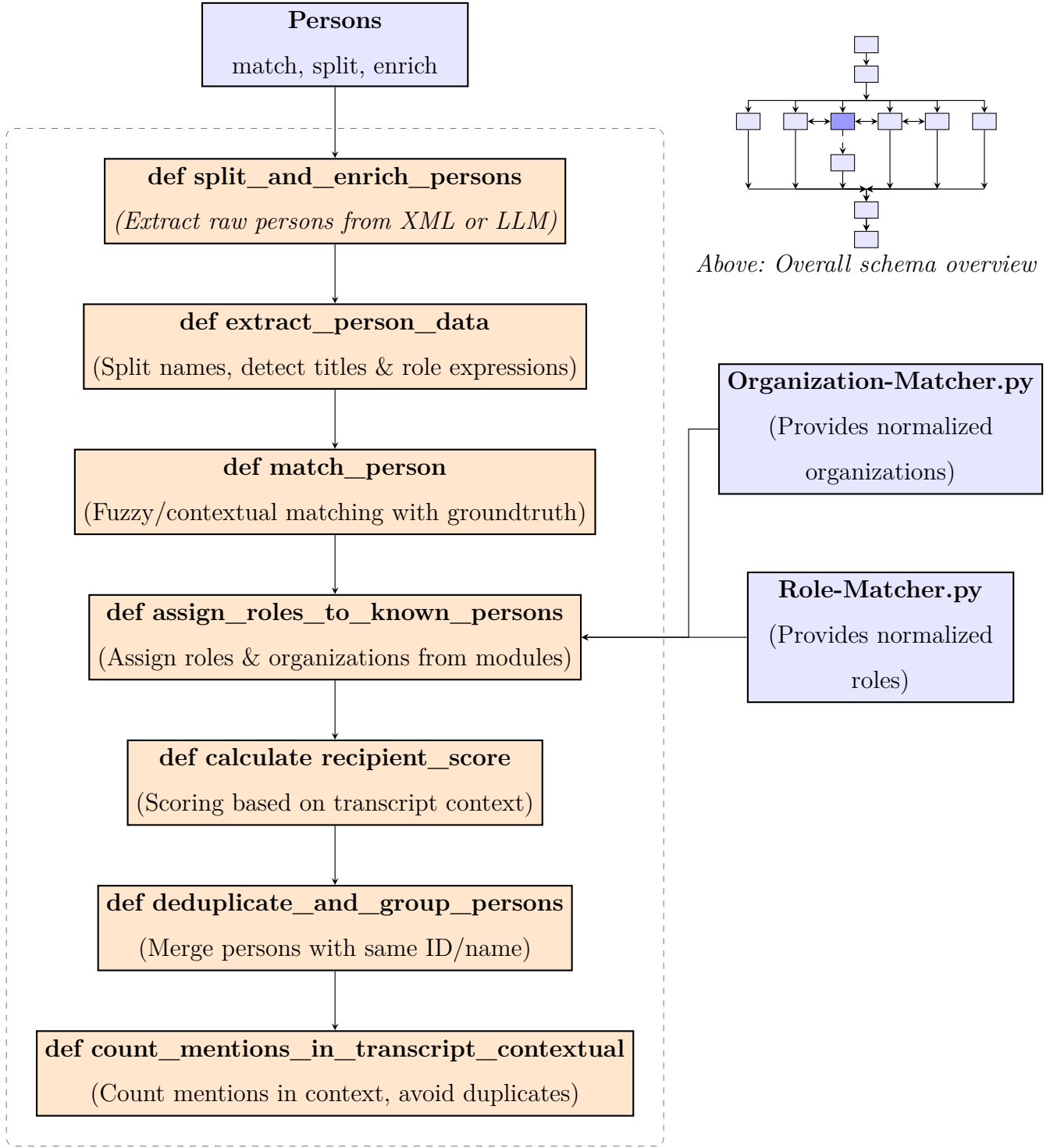
Dein

Carl

Abbildung 7: Tagging von [Abbildung 6](#)

# Diagram english





Above: Overall schema overview

Top: process in person\_matcher.py

Right: input from role\_matcher.py and organization\_matcher.py

## Gründe für den Wechsel zu Nodegoat

### Nodegoat Modellierung

## Netzwerkanalyse als Methode

### Theoretischer Hintergrund der Netzwerkanalyse

### Ziele der Netzwerkanalyse im Kontext der Quellen

### Technische Umsetzung (Tools, Datenbankstruktur)

## Normalisierung der Dateien — von PDF zu JPEG

```
1 import os
2 import fitz  # PyMuPDF
3
4 def convert_pdf_to_jpg(src_folder, dest_folder):
5     # Überprüfen, ob der Zielordner existiert, und ihn ggf. erstellen
6     if not os.path.exists(dest_folder):
7         os.makedirs(dest_folder)
8
9     # Durchgehen durch alle Dateien im Quellordner
10    for root, dirs, files in os.walk(src_folder):
11        for file in files:
12            # Überprüfen, ob die Datei eine PDF-Datei ist
13            if file.lower().endswith(".pdf"):
14                # Vollständigen Pfad zur PDF-Datei erstellen
15                pdf_path = os.path.join(root, file)
16                # PDF-Datei öffnen
17                doc = fitz.open(pdf_path)
18                # Durch alle Seiten der PDF-Datei gehen
19                for page_num in range(len(doc)):
20                    page = doc[page_num]
21                    # Seite in ein PiXMap-Objekt umwandeln (für die Konvertierung in
22                    ↪ JPG)
23                    pix = page.get_pixmap()
24                    # Dateinamen ohne Dateiendung extrahieren
25                    filename_without_extension = os.path.splitext(file)[0]
26                    # Ausgabedateinamen erstellen mit führenden Nullen für die
27                    # Seitennummer
28                    output_filename = f"{filename_without_extension}_S{page_num +
29                    ↪ 1:03d}.jpg"
30
31                    # Vollständigen Pfad zur Ausgabedatei erstellen
32                    output_path = os.path.join(dest_folder, output_filename)
33                    # Bild speichern
```

```

33         pix.save(output_path)
34         # PDF-Datei schließen
35         doc.close()
36
37         # Erfolgsmeldung ausgeben
38         print(f"{file} wurde erfolgreich umgewandelt und gespeichert
39         in {dest_folder}")
40
41     # Pfade zu den Ordnern mit den PDF-Dateien (Quelle) und den JPG-Dateien (Ziel)
42     src_folder = r"/Users/svenburkhardt/Documents/D_Murger_Männer_Chor_Forschung/Scan_Mä_
↳ nnerchor/Männerchor_Akten_1925-1945/Scan_Männerchor_PDF"
43     dest_folder = r"/Users/svenburkhardt/Documents/D_Murger_Männer_Chor_Forschung/Master_
↳ arbeit/JPEG_Akten_Scans"
44
45
46     # Funktion aufrufen, um die Konvertierung durchzuführen
47     convert_pdf_to_jpg(src_folder, dest_folder)
48

```

# Aufbau der Datenbank

## Konzeption der Datenmodellierung

Eigene Ontologie im Vergleich zu bestehenden Standards

Verknüpfung von Personen, Orten und Ereignissen

## Implementierung der Datenbank

Datenbankdesign

Herausforderungen bei der Datenaufnahme

Verknüpfung mit externen Quellen (z.B. Wikidata)

# Analyse der Netzwerke

## Soziale Netzwerke des Vereinslebens

Verbindungen zwischen Mitgliedern

Kooperationen mit anderen Vereinen

## Politische Netzwerke und deren Veränderungen

Einfluss der NS-Diktatur auf die Netzwerke

Feldpostkarten als Quelle für militärische Netzwerke

## Geografische Ausdehnung der Netzwerke

Einsatzorte der Chormitglieder während des Krieges

Lokale und überregionale Verbindungen

## Diskussion der Ergebnisse

Sichtbarmachung der Netzwerke durch Nodegoat und Netzwerkanalyse

Gibt es Veränderungen der Netzwerke im historischen Kontext?



# Bibliographie

Referenzen, die noch nachzuschauen sind:

– Akten\_Gesamtübersicht.csv im Anhang

## References

„78th Sturm-Division (Wehrmacht)“. 2024. Unter Mitarbeit von Sven Burkhardt, (Zugriff am besucht am 12. März 2025)

. Besucht am 12. März 2025. <https://www.wikidata.org/wiki/Q125489568>.

Burkhardt, Sven. 2022. „ArcGIS StoryMaps“. ArcGIS StoryMaps, (Zugriff am besucht am 12. März 2025)

. Besucht am 12. März 2025. <https://storymaps.arcgis.com>.

„DRK Suchdienst | Suche per Feldpostnummer“. DRK Suchdienst; Suche per Feldpostnummer. 2025. Unter Mitarbeit von Christian Reuter, (Zugriff am besucht am 12. März 2025)

. Besucht am 12. März 2025. <https://vbl.drk-suchdienst.online/Feldpostnummer/FPN.aspx>.

„Forum Geschichte der Wehrmacht“. Forum Geschichte der Wehrmacht. o. Dat. Unter Mitarbeit von Dieter Hermans, (Zugriff am besucht am 12. März 2025)

. Forum. Besucht am 12. März 2025. <https://www.forum-der-wehrmacht.de/>.

Gemeinde Murg, Hrsg. *Geschichte Gemeinde Murg*

. Besucht am 29. Juni 2025. <https://www.murg.de/seite/33378/geschichte.html>.

Hartmann, Christian. 2010. *Wehrmacht im Ostkrieg - Front und militärisches Hinterland 1941/42*. 2. Auflage. Bd. 75. Quellen und Darstellungen zur Zeitgeschichte Herausgegeben vom Institut für Zeitgeschichte. München: R. Oldenbourg Verlag

.

Hollmann, Prof. Dr. Michael. 2025. „Freiburg“. Bundesarchiv Freiburg im Breisgau (Abteilung Militärarchiv), (Zugriff am besucht am 12. März 2025)

. Besucht am 12. März 2025. <https://www.bundesarchiv.de/das-bundesarchiv/standorte/freiburg/>.

„Lexikon der Wehrmacht“. o. Dat. Unter Mitarbeit von Andreas Altenburger, (Zugriff am besucht am 12. März 2025)

. Besucht am 12. März 2025. <http://www.lexikonderwehrmacht.de/>.

Rass, Christoph und René Rohrkamp. 2009. *Deutsche Soldaten 1939-1945 Handbuch einer biographischen Datenbank zu Mannschaften und Unteroffizieren von Heer, Luftwaffe und Waffen-SS*. Aachen

.

Tessin, Georg. 1977. *Verbände und Truppen der deutschen Wehrmacht und Waffen-SS im Zweiten Weltkrieg 1939-1945*. Bd. Band 1 - Die Waffengattungen — Gesamtübersicht. Osnabrück: HIBLIO Verlag

.