



# Von Papier zur digitalen Netzwerkanalyse

---

Digitalisierung, Modellierung und Untersuchung  
historischer Vereinsakten  
mit Machine Learning und Nodegoat

Sven Burkhardt

 0009-0001-4954-4426

 17-056-912

 17-01-2025




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 der Weimarer Republik bis zum Ende des Zweiten Weltkrieges. 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
<b>Historischer Kontext</b>	<b>2</b>
Historische Einordnung des Zeitraums . . . . .	2
Historische Einordnung des Vereins . . . . .	2
Der Männerchor während des Zweiten Weltkriegs . . . . .	2
Politische Entwicklungen und ihre Auswirkungen auf das Vereinsleben . . . . .	2
<b>Quellenbeschreibung und Korpusaufbau</b>	<b>3</b>
Beschreibung des Archivbestands . . . . .	3
<b>Methodischer Zugang</b>	<b>4</b>
Digitale Erfassung und Strukturierung der Quellen . . . . .	4
Gliederung in Akten . . . . .	4
Digitalisierung und Transkription . . . . .	4
Tagging in Transkribus . . . . .	4
Strukturelle Tags . . . . .	4
Inhaltliche Tags . . . . .	5
Digitalisierungsprozess und Herausforderungen . . . . .	7
Wechsel von Linked Open Data (LOD zu Nodegoat) . . . . .	13
Definition und Nutzen von LOD . . . . .	13
Aufbau der LOD Ontologie . . . . .	13
Gründe für den Wechsel zu Nodegoat . . . . .	13
Nodegoat Modellierung . . . . .	13
Netzwerkanalyse als Methode . . . . .	13
Theoretischer Hintergrund der Netzwerkanalyse . . . . .	13
Ziele der Netzwerkanalyse im Kontext der Quellen . . . . .	13
Technische Umsetzung (Tools, Datenbankstruktur) . . . . .	13
Normalisierung der Dateien — von PDF zu JPEG . . . . .	13
<b>Aufbau der Datenbank</b>	<b>15</b>
Konzeption der Datenmodellierung . . . . .	15
Eigene Ontologie im Vergleich zu bestehenden Standards . . . . .	15
Verknüpfung von Personen, Orten und Ereignissen . . . . .	15
Implementierung der Datenbank . . . . .	15
Datenbankdesign . . . . .	15
Herausforderungen bei der Datenaufnahme . . . . .	15
Verknüpfung mit externen Quellen (z.B. Wikidata) . . . . .	15
<b>Analyse der Netzwerke</b>	<b>16</b>
Soziale Netzwerke des Vereinslebens . . . . .	16
Verbindungen zwischen Mitgliedern . . . . .	16
Kooperationen mit anderen Vereinen . . . . .	16
Politische Netzwerke und deren Veränderungen . . . . .	16
Einfluss der NS-Diktatur auf die Netzwerke . . . . .	16

Feldpostkarten als Quelle für militärische Netzwerke . . . . .	16
Geografische Ausdehnung der Netzwerke . . . . .	16
Einsatzorte der Chormitglieder während des Krieges . . . . .	16
Lokale und überregionale Verbindungen . . . . .	16
<b>Diskussion der Ergebnisse</b>	<b>17</b>
Sichtbarmachung der Netzwerke durch Nodegoat und Netzwerkanalyse . . . . .	17
Gibt es Veränderungen der Netzwerke im historischen Kontext? . . . . .	17
<b>Fazit und Ausblick</b>	<b>18</b>
Zusammenfassung der zentralen Erkenntnisse . . . . .	18
Methodische Herausforderungen und Lösungen . . . . .	18
Ausblick auf zukünftige Forschung und mögliche Erweiterungen der Datenbank .	18
<b>Bibliographie</b>	<b>19</b>

**Einleitung**

**Ziel und Relevanz der Arbeit**

**Forschungsstand und Forschungslücke**

**Formulierung der Forschungsfrage**

**Aufbau der Arbeit**

# Historischer Kontext

Historische Einordnung des Zeitraums

Historische Einordnung des Vereins

Der Männerchor während des Zweiten Weltkriegs

Politische Entwicklungen und ihre Auswirkungen auf das Vereinsleben

# Quellenbeschreibung und Korpusaufbau

## Beschreibung des Archivbestands

## Methodischer Zugang

### Digitale Erfassung und Strukturierung der Quellen

#### Gliederung in Akten

#### 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:

#### 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.), Murgtal-schule, 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. In insgesamt 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 beigegefügt. Die so erstellten 70 Akten ergaben 158 Seiten zu insgesamt 22.155 Wörtern Groundtruth, womit dann ein eigenes Transkribus Modell ([ModelID: 287793](#)) erstellt wurde. Es erreichte eine Accuracy (CER) von 6,58%. Später wurden die verbleibenden 80 Akten nur noch mit diesem Modell transkribiert.

Beispiel für handschriftlichen Text erkannt von Transkribus in Akte\_076:

Krieg, 15. Aug. 41.

Mein lieber Alois!

Deine lange Briefe ab und an dem Maimacher wieder einmal eine Liederung zu wissen. und kam mir die ganze Gelegenheit zu passen.

Ich habe nunmehr den Namen der Maimacher Teilstadt um den Titel der Liederung zu erfüllen, wo sie zum Abfinden der Arbeit gehen "wie Windsturm" über oft in Paris - meiste beifolgt, keine Antwort. Vielleicht gelingt es dir diesen Titel zu erfüllen.

Die Liederung ist die Liederung  
"Der alte Lied" von dem. Kirsch

Es wurde 1928 um 10. Teilst. Liederung. Post von Liederungsbuch in die Liederung und wurde in einem großen Brieffall.

Es ist schon der Richtig zu finden.

Als Alois: zuerst der Teilstadt Liederung, wenn der Liederung und wenn Liederung unmöglich, dann sein Wahl.

Mit sehr. Grüssen

Ein

Carl.

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 Iitel 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

ChatGPT produziert daraus:



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

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:

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 Jahhr an den  
Männerchor 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 herzl. Grüßen

Dein

Carl



# Wechsel von Linked Open Data (LOD zu Nodegoat)

## Definition und Nutzen von LOD

## Aufbau der LOD Ontologie

## 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"
```

```

28
29
30         # Vollständigen Pfad zur Ausgabedatei erstellen
31         output_path = os.path.join(dest_folder, output_filename)
32         # 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?

## Fazit und Ausblick

Zusammenfassung der zentralen Erkenntnisse

Methodische Herausforderungen und Lösungen

Ausblick auf zukünftige Forschung und mögliche Erweiterungen  
der Datenbank

## Bibliographie