

Projektvorstellung aus:  
Zeitschrift für digitale Geisteswissenschaften, Heft 11 (2026)

Titel:  
OHTM-Dashboard. Interaktives visuelles Topic-Modeling-Analyse-Tool

Autor\*in:  
Philipp Bayerschmidt


Kontakt: [philipp.bayerschmidt@fernuni-hagen.de](mailto:philipp.bayerschmidt@fernuni-hagen.de)  
Institution: FernUniversität in Hagen  
GND: [139340135X](#) ORCID: [0009-0002-2743-2455](#)  
Contribution (CRediT): [Conceptualization](#) | [Formal analysis](#) | [Investigation](#) | [Software](#) | [Writing – original draft](#)

---

DOI des Beitrags:  
[10.17175/2026\\_009](https://doi.org/10.17175/2026_009)

Nachweis im OPAC der Herzog August Bibliothek:  
[1971798894](#)

Erstveröffentlichung:  
01.07.2026

Lizenz:  
Sofern nicht anders angegeben 

Letzte Überprüfung aller Verweise:  
02.06.2026

Format:  
PDF ohne Paginierung, Lesefassung

GND-Verschlagwortung:  
[Oral History](#) | [Topic Modeling](#) | [Visual Analytics](#) | [Softwareentwicklung](#)

Empfohlene Zitierweise:  
Philipp Bayerschmidt: OHTM-Dashboard. Interaktives visuelles Topic-Modeling-Analyse-Tool. In: Zeitschrift für digitale Geisteswissenschaften 11 (2026). 01.07.2026. HTML / XML / PDF. DOI: [10.17175/2026\\_009](https://doi.org/10.17175/2026_009)

Philipp Bayerschmidt

# OHTM-Dashboard. Interaktives visuelles Topic-Modeling-Analyse-Tool

## Abstract

Im *Topic Modeling* gestaltet sich die Analyse der statistischen Ergebnisse als umständlich und bedarf häufig einer visuellen Aufarbeitung. Das im Rahmen des DFG-geförderten Projekts ›Oral-History.Digital‹ entwickelte Tool *Oral History Topic Modeling Dashboard (OHTM-Dashboard)* bietet einen Lösungsansatz und ist auf die spezifischen Anforderungen lebensgeschichtlicher Interviews optimiert. Nach einer methodischen Einführung in das *LDA*-Verfahren und einem vergleichenden Überblick über bestehende Analyse-Tools werden die Dashboard-Funktionen präsentiert. Anhand einer Case Study zur Analyse von Heimat- und Beheimatungsprozessen in einem Korpus von 991 Interviews aus sieben verschiedenen Archiven wird der Einsatz von OHTM-Dashboard gezeigt.

In *topic modeling*, analyzing statistical results can be complex and often requires visual interpretation. The *Oral History Topic Modeling Dashboard (OHTM Dashboard)*, a tool developed as part of the DFG-funded project ›Oral-History.Digital‹, offers a solution and is optimized for the specific requirements of life history interviews. Following a methodological introduction to the *LDA* procedure and a comparative overview of existing analysis tools, the dashboard functions will be presented. The use of OHTM-Dashboard will be demonstrated through a case study analyzing processes of home and settlement in a corpus of 991 interviews from seven different archives.

## 1. Einleitung

Die wachsende Zugänglichkeit von Oral-History-Interviewbeständen durch die fortschreitende Digitalisierung, insbesondere durch die Entstehung des Online-Archivs *Oral-History.Digital (OH.D)*, führt zu der Frage, wie die Vielzahl von Interviews zukünftig analysiert und erforscht werden kann. Welche Möglichkeiten – neben den Standardlösungen wie Volltextsuche oder Metadatenfiltern – gibt es, um tausend Interviews aus verschiedenen Archiven und unterschiedlichen thematischen Forschungsprojekten perspektivisch in einer Sekundäranalyse sammlungsübergreifend zu untersuchen? Um dieser Frage nachzugehen, wurde ein Forschungskorpus mit insgesamt 991 Interviews aus sieben verschiedenen Archiven zusammengestellt: aus dem *Archiv Deutsches Gedächtnis (ADG)* des *Instituts für Geschichte und Biographie der FernUniversität in Hagen*, dem *Archiv Zwangsarbeit 1939–1945 (ZWA)* und *Colonia Dignidad*. Ein *chilenisch-deutsches Oral-History-Archiv (CDG)* der *Freien Universität Berlin*, der *Interviewsammlung der Werkstatt der Erinnerung (WdE)* der *Forschungsstelle für Zeitgeschichte in Hamburg*, dem *Zeitzeugenarchiv des Dokumentationszentrum Flucht, Vertreibung, Versöhnung (FVV)*, der *Sammlung Erzählte Lebensgeschichte des Museums Friedland (MFL)* und einer *Sammlung des Hannah-Arendt-Instituts für Totalitarismusforschung (HAIT)* zu Kindergarten und Schule in der DDR. Ausschlaggebend für die Zusammenstellung des Korpus war zum einen die Abbildung der thematischen Breite des Online-Archivs und zum anderen die Voraussetzung, dass die Transkripte vollständig digitalisiert vorlagen. Da ein solches Korpus zu umfangreich für eine manuelle Analyse ist, wurde die Methode des *Topic Modelings*<sup>1</sup>, aufbauend auf gemeinsamen Vorarbeiten mit Dennis Möbus, Tobias Hodel und Ina Serif,<sup>2</sup> gewählt.<sup>3</sup> *Topic Modeling* ist ein probabilistisches Verfahren aus dem Bereich des *Text Mining*, einem Unterbereich des *Natural Language Processing*, bei dem das gemeinsame Vorkommen von Wörtern innerhalb eines Textkorpus auf Grundlage von statistischen Verteilungen berechnet und zu verschiedenen Gruppen, sogenannten *Topics*, zugeordnet wird. Die *Topics* können thematisch interpretiert und in die jeweiligen Dokumente zurückverfolgt werden. Diese Zurückführung auf die jeweiligen Dokumente bzw. Transkripte ist ein zentraler Bestandteil der Analyse. In der Kombination von *distant* und *close reading* können Erkenntnisse zum Vorschein kommen, die über eine rein manuelle Analyse nicht

[1]

<sup>1</sup> Vgl. Blei et al. 2003, S. 993–1022.

<sup>2</sup> Vgl. Möbus 2020, S. 162–180; Hodel et al. 2022; Franken / Möbus 2024.

<sup>3</sup> Vgl. Bayerschmidt 2026.

gefunden worden wären.<sup>4</sup> Dieser Wechsel der Perspektive, auch bekannt als *scalable reading*,<sup>5</sup> ist kein einmaliges Vorgehen, sondern ein stetiges Wechselspiel, bei dem die Ergebnisse auf verschiedenen Ebenen und aus verschiedenen Perspektiven miteinander in Bezug gesetzt werden. Da die manuelle Auswertung der Topic-Modeling-Ergebnisse und der Wechsel der verschiedenen Ebenen sehr zeitaufwendig und vor allem sehr unübersichtlich sind, wurde ein eigenes interaktives visuelles Topic-Modeling-Analyse-Tool entwickelt, das *Oral History Topic Modeling Dashboard (OHTM-Dashboard)*.<sup>6</sup>

## 2. Topic Modeling

Es gibt eine Vielzahl unterschiedlicher Algorithmen für Topic Modeling. In einer Studie haben Ina Serif, Tobias Hodel und Dennis Möbus herausgearbeitet, dass das *Latent Dirichlet Allocation-Verfahren (LDA)*<sup>7</sup> für die Arbeit mit lebensgeschichtlichen Interviews am besten funktioniert.<sup>8</sup> LDA geht von der Grundannahme aus, dass jedes Dokument eine Zusammensetzung verschiedener Themen ist. Das Verfahren berechnet die statistische Verteilung und Nähe der Wörter zueinander und teilt diese einzelnen Topics zu, die Anzahl der Topics ist dabei frei wählbar. Als Ergebnis erhält man die Topic-Wort-Verteilung und die Topic-Dokument-Verteilung. Die Topic-Wort-Verteilung gibt die Wörter der jeweiligen Topics, absteigend nach ihrer Gewichtung, als Wortliste an. Die Topic-Dokument-Verteilung enthält die Wahrscheinlichkeiten, die jedes Topic innerhalb eines Dokuments besitzt. [2]

Topic-Nummer	Ersten 20 Wörter des Topics
44	deutsch, polen, deutschland, polnisch, sprache, russisch, pole, lernen, land, deutsche, familie, russland, friedland, türkei, stadt, russe, heimat, ausländer, türkisch, grenze,

Tab. 1: Topic-Wort-Verteilung des Topics 44.

Topic	Topic-Gewichtung
44	0.5990997305708491
94	0.21392577268082863
0	0.04873578232889533
87	0.02992986020180754
50	0.020996276856984443

Tab. 2: Beispielhafte Topic-Dokument-Verteilung.

Es ist wichtig anzumerken, dass die Wortlisten der Topics alleine auf der statistischen Verteilung der Wörter innerhalb des Forschungskorpus gebildet werden und keine externen Informationen für die Berechnung verwendet werden. Die Wortlisten besitzen aber dennoch semantische Gemeinsamkeiten, die aus dem gemeinsamen Auftreten innerhalb von Sätzen oder Textabschnitten resultieren. Die Themen ergeben sich somit aus dem thematischen Kontext, in dem die Wörter zusammen verwendet werden. Diese können sich auf ein oder mehrere Themen beziehen, abhängig vom Kontext des Korpus. Um diese inhaltlichen Gemeinsamkeiten in Form von Themen greifen zu können, müssen die Ergebnisse durch die Forschenden interpretiert werden. Nur so können die Topics thematisch genutzt und in die Textabschnitte zurückverfolgt werden. [3]

<sup>4</sup> Vgl. Jänicke et al. 2017, S. 226–250; Krautter 2024, S. 261–286.

<sup>5</sup> Vgl. Zepezauer-Wachauer 2021.

<sup>6</sup> Vgl. Bayerschmidt 2025–2026; das Dashboard kann online auf der Archivseite des ADG frei zugänglich ausprobiert werden: <https://ohtm-dash.oral-history.digital/>.

<sup>7</sup> Vgl. Blei et al. 2003.

<sup>8</sup> Vgl. Hodel et al. 2022.

Bevor die Topics berechnet werden können, muss das zu untersuchende Textkorpus vorbereitet werden. Im *Preprocessing* können verschiedene Methoden gewählt und miteinander kombiniert werden, um die Ergebnisse des Topic Modelings für die Analyse zu verbessern.<sup>9</sup> [4]

1. Die Entfernung von *Stopwords*, also Wörtern, die sehr häufig vorkommen und deshalb die Ergebnisse verrauschen können. Dies können beispielsweise Pronomen, Konjunktionen, Präpositionen etc. sein. Die Auswahl der zu entfernenden Stopwords ist stark von der jeweiligen Quellenart und der Forschungsfrage abhängig.
2. Die Bereinigung von Satzzeichen, Paratext sowie Timecodes und Sprecher\*innenkürzeln.
3. Die Lemmatisierung von Wörtern, also die Rückführung von Wörtern auf ihre Grundform.
4. Die Unterteilung der Dokumente und ihrer Sätze in einzelne Wörter bzw. Wortlisten, die sogenannte Tokenisierung, damit die Berechnung auf Grundlage der einzelnen Wörter erfolgen kann.

Aus dem Originaltext »Und das war dann mein großes Glück! (lacht) Habe noch nie soviel Geld, wie ich da verdient habe.« wird nach dem Preprocessing z. B. dieser verarbeitete Text: »großes, glück, geld, verdienen«. [5]

Diese vorverarbeiteten Wortlisten sind, je nach Länge der einzelnen Dokumente des Korpus, die Grundlage des Topic Modelings. Auf ihnen werden die statistische Verteilung der Wörter und damit die Topics berechnet. Im Fall der lebensgeschichtlichen Interviews sind die Transkripte von einem Umfang zwischen 60 und 100 Seiten deutlich zu lang für eine direkte Berechnung, weshalb sie in kleinere Textdokumente, sogenannte *Chunks*, unterteilt werden müssen. Für Interviews hat sich in gemeinsamen Vorarbeiten eine Länge von 500 Wörtern pro Chunk bzw. 100 Wörtern pro Chunk nach der Stopwords-Entfernung als geeignet erwiesen.<sup>10</sup> Lebensgeschichtliche Interviews bestehen aus einer freien Erzählung, die häufig Themenwechsel oder andere thematische Einschübe enthalten, die durch das *Chunking* in der Topic-Verteilung berücksichtigt werden können. [6]

## 2.1 Analyseebenen des Topic-Modelings

Die beiden ersten Analyseebenen, die Topic-Wort-Verteilung und die Topic-Dokument-Verteilung, bilden die Grundlage für das distant reading. Für das close reading ist es wichtig, die Topic-Dokument-Verteilung auf die nicht verarbeiteten Texte zurückzuführen. Wurden die Dokumente nicht gechunkt, ist diese Zurückführung leicht umsetzbar. Bei gechunkten Dokumenten müssen die Dateien aber bereits vor der Topic-Berechnung so aufbereitet worden sein, dass die Ergebnisse der Chunks auf die jeweiligen Textstellen im Dokument zurückverfolgt werden können. Wäre das Chunking ein einmaliger Vorgang, der nach festen Regeln erfolgt, könnten die jeweiligen Dokumente, in diesem Fall Interviews, zuvor in kleinere Teildokumente unterteilt werden und erst dann dem Korpus hinzugefügt werden. Allerdings ist der Vorgang nicht einheitlich, sondern stark von den Inhalten und der Struktur der Dokumente abhängig und auch stark vom jeweiligen Forschungsinteresse geprägt.<sup>11</sup> Somit stellt das Chunking einen dynamischen Prozess dar, der erst durch Evaluation der Ergebnisse festgelegt werden kann. Deshalb bietet die OHTM-Pipeline (vgl. Kapitel 4) und das OHTM-Dashboard ein variables System, das unabhängig von der gewählten Chunklänge die Ergebnisse ohne eigene Vorarbeiten auf die Textstellen und auch auf das Originaldokument zurückverfolgen kann. Bei gechunkten Dokumenten erhält man durch die Addition der Topic-Dokument-Verteilung aller zum Dokument gehörigen Teildokumente die durchschnittliche Topic-Verteilung für das gesamte Dokument bzw. in diesem Fall für das gesamte Interview. Addiert man die Topic-Dokument-Verteilung aller Dokumente zusammen, erhält man die Korpusebene, die Topic-Verteilung für das Forschungskorpus. Wenn das Korpus nicht nur [7]

<sup>9</sup> Eine detaillierte Ausführung des Preprocessings und seiner Einflüsse findet sich in Möbus et al. 2026.

<sup>10</sup> Vgl. Möbus et al. 2026.

<sup>11</sup> Vgl. Möbus et al. 2026, S. 21–27.

aus einer Sammlung besteht, sondern wie in diesem Fall aus verschiedenen Archiven, ist es wichtig, die Korpusebene für die verschiedenen Teilkorpora bilden zu können. Insgesamt ergeben sich bei der Arbeit mit Topic Modeling folgende Analyseebenen:

- Topic-Wort-Verteilung (Wortliste)
- Topic-Chunk-Verteilung
- Topic-Dokument-Verteilung → Dokumentebene
- unverarbeitetes Dokument → Dokumentansicht
- Teilkorpusebene
- Korpusebene

All diese Analyseebenen können für eine Auswertung der Ergebnisse kombiniert und miteinander in Bezug gesetzt werden, je nach Interesse. [8]

### 3. Bestehende Topic-Modeling-Analyse-Tools

Obwohl die statistischen Ergebnisse des Topic Modelings sehr gute Einblicke in den Aufbau und mögliche Inhalte von Dokumenten liefern, sind diese an sich jedoch schwer zu erschließen und zu verstehen.<sup>12</sup> Deshalb haben sich schon früh Visualisierungen für die Interpretation als mögliche Lösung herauskristallisiert und sind bis heute ein fester Bestandteil des Topic Modelings.<sup>13</sup> Im Folgenden wird ein kurzer Überblick über bereits bestehende Lösungen gegeben und aufgezeigt, welche Schwachstellen in Bezug auf die Darstellung der verschiedenen Analyseebenen bestehen. [9]

#### 3.1 LDAvis und Topic Cloud

Angebote wie ›LDAvis<sup>14</sup> (Abbildung 1) oder die ›Topic Cloud<sup>15</sup> (Abbildung 2) bieten eine informative Darstellung der Topic-Wort-Verteilung und zeigen die Verhältnisse der Topics und der Wörter untereinander sowie zu anderen Topics an. Zusätzlich wird auch die Topic-Verteilung für das gesamte Korpus visualisiert. [10]

---

<sup>12</sup> Vgl. beispielsweise Chaney / Blei 2012, S. 1.

<sup>13</sup> Vgl. Blei 2012, S. 84; Sievert / Shirley 2014; Chuang et al. 2012; Liu et al. 2021.

<sup>14</sup> Vgl. Sievert / Shirley 2014.

<sup>15</sup> Vgl. Rönnqvist et al. 2014.



### 3.2 Uncertainty-Aware Topic Modeling Visualization

Valerie Müller, Christian Sieg und Lars Linsen haben 2021 ein interaktives Tool publiziert, welches sie für die Exploration von Unsicherheiten im Topic Modeling nutzen, darunter verstehen sie die Abweichungen und Unterschiede verschiedener Modelle mit unterschiedlichen Parametern.<sup>16</sup> Für die Topic-Dokument-Verteilung wurde ein Balkendiagramm gewählt. [12]

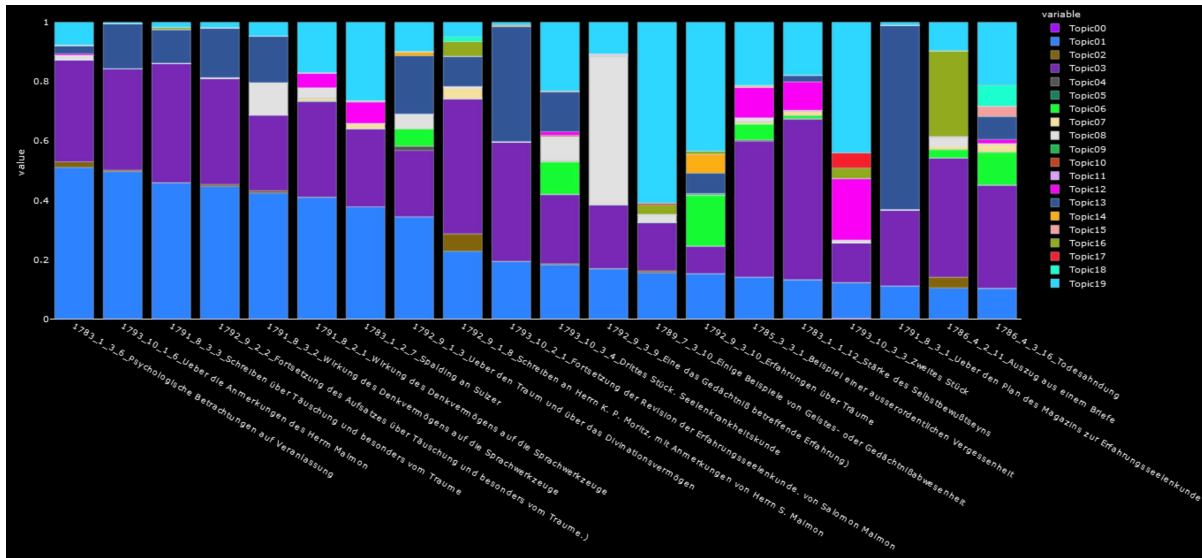


Abb. 3: Topic-Dokument-Verteilung. [Aus: Müller et al. 2021, S. 4]

Die qualitative Analyse erfolgt in einer Dokumentansicht mit dem unverarbeiteten Text sowie den farblichen Markierungen der Wörter zu ihren Topics. Mehrere Fenster bieten weitere Informationen.<sup>17</sup> Leider ist das Tool nicht *open source* zugänglich und kann daher nicht weiterentwickelt werden, weshalb das Tool hier nicht weiter beachtet wird. [13]



Abb. 4: Dokumentenansicht für das close reading. [Aus: Müller et al. 2021, S. 5]

<sup>16</sup> Vgl. Müller et al. 2021.

<sup>17</sup> Vgl. Müller et al. 2021, S. 4-5.

### 3.3 DARIAH Topics Explorer

Im deutschsprachigen Raum ist der **DARIAH Topics Explorer** bekannt, eine kompakte Softwarelösung, die die einfach bedienbare Berechnung eines Topic Modells und deren visuelle Auswertung bietet.<sup>18</sup> Die Ergebnisse können auf mehreren Ebenen analysiert und teilweise interaktiv verknüpft werden. Zunächst kann die Topic-Wort-Verteilung angezeigt werden, und über einen Klick auf ein Topic gelangt man zu einer Topic-Dokument-Verteilung für das gewünschte Topic.

[14]

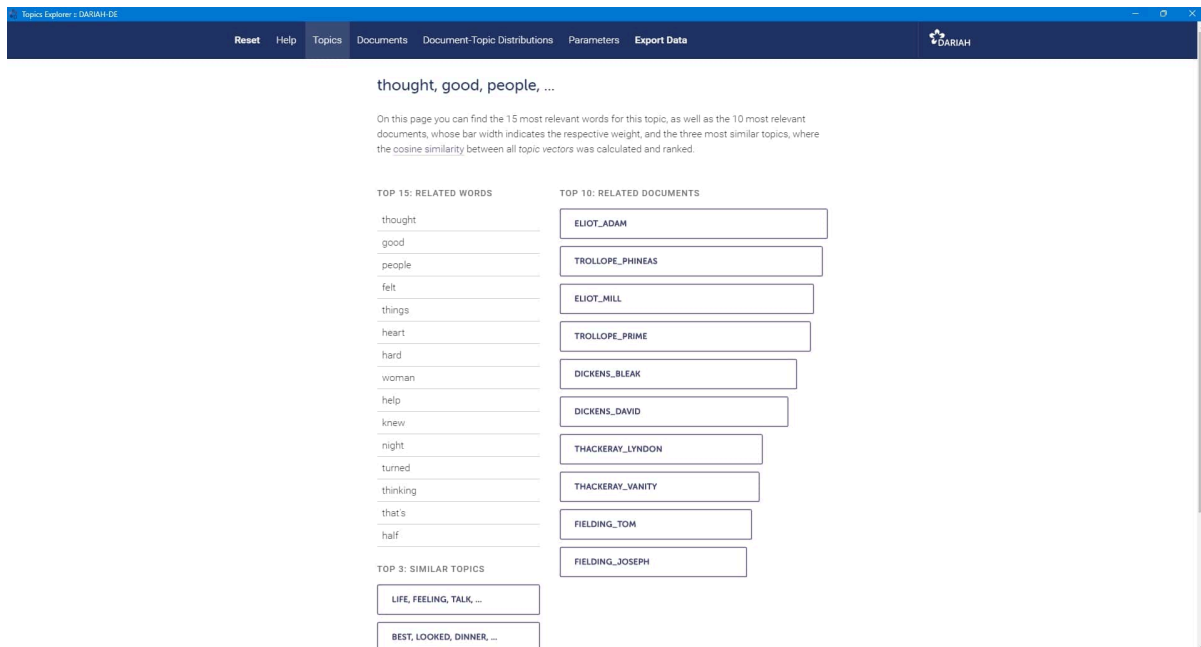


Abb. 5: Topic-Dokument-Verteilung für das ausgewählte Topic im DARIAH Topics Explorer [Screenshot: Philipp Bayerschmidt 2026]

Durch einen Klick auf das jeweilige Dokument gelangt man zu einer detaillierteren Topic-Dokument-Verteilung für das ausgewählte Dokument samt Rückbezug auf die unverarbeiteten Texte. Jedoch ist die Topic-Übersicht am Seitenanfang wenig intuitiv und liefert nur mit geschicktem Einsatz des Mauszeigers weiterführende Informationen.

[15]

<sup>18</sup> Vgl. Simmler et al. 2019.

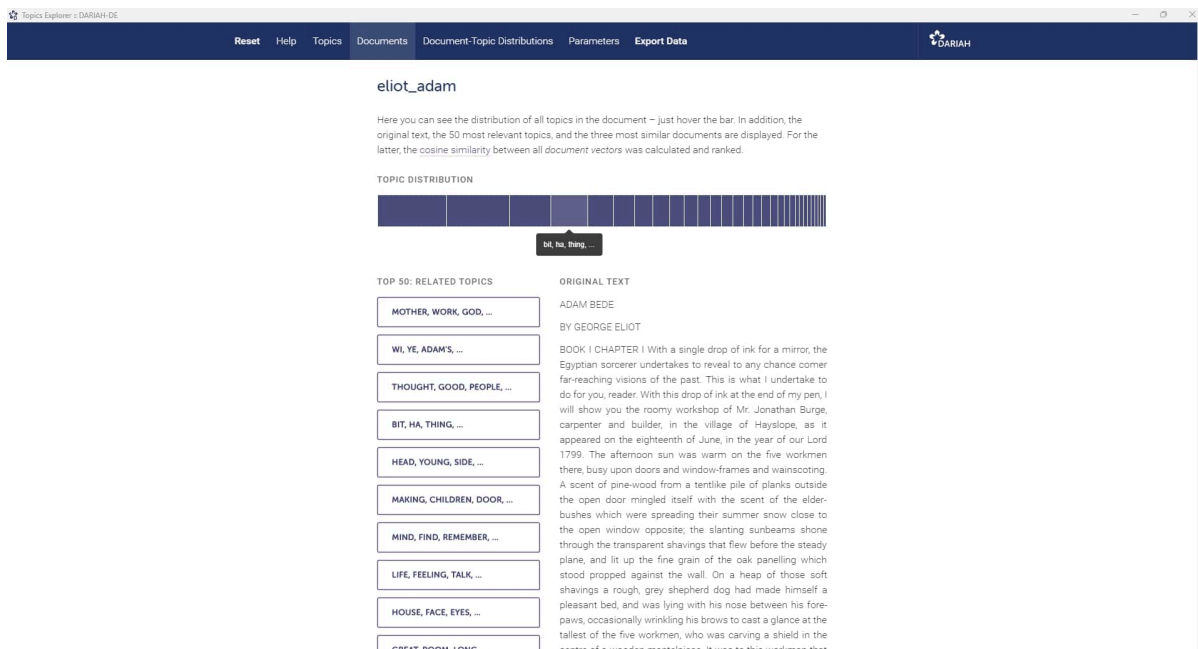


Abb. 6: Detaillierte Topic-Dokument-Verteilung für ein Dokument im DARIAH Topics Explorer. [Screenshot: Philipp Bayerschmidt 2026]

Zusätzlich gibt es auch eine gesamte Topic-Dokument-Verteilung in Form einer *Heatmap*, allerdings ohne interaktive Verknüpfung zurück in die jeweiligen Dokumente.

[16]



Abb. 7: Topic-Dokument-Verteilung des DARIAH Topics Explorers. [Screenshot: Philipp Bayerschmidt 2026]

Mehrere Anzeigen gleichzeitig darzustellen, ist nicht möglich, da sich bei einer Verlinkung der Ebenen immer die neue Ebene öffnet und die vorherige Ebene schließt. Die gleichzeitige Darstellung mehrerer Ebenen würde eine bessere Verknüpfung der Ergebnisse ermöglichen und das scalable reading ungemein erleichtern. Die Berechnung der Topic-Verteilung erfolgt lediglich auf Grundlage von TXT-Dateien oder XML-Dateien, zudem fehlt die Möglichkeit, die Texte zu chunken oder Teilkorpora zu bilden. Ein *Parametertuning*, das

[17]





Abb. 9: Analyseansicht des TwiC. [Aus: Armoza 2017]

Die Dokumente können in der unverarbeiteten Form angezeigt werden, wobei die Wörter der Topics und die Stopwords farblich markiert werden. In der Ansicht werden in Teilfenstern weitere wichtige Informationen angezeigt, wie die Korpusebene (links oben), Topic-Dokument-Verteilung (rechts oben), Topic-Wort-Verteilung mit Textvorschau (links unten) und die Dokumentenansicht (rechts unten). Diese sind interaktiv miteinander verbunden. Die Anwendung ist aber, besonders in der visuellen Repräsentation, auf Gedichte und deren Länge optimiert. So wird in der Topic-Dokument-Verteilung keine vollständige Liste angezeigt, sondern eine abstrakte Vorschau des Dokuments, in der die ersten zehn Zeilen als Rechtecke abgebildet sind, farblich markiert nach dem entsprechenden Topic. Die Topics werden als Ringe dargestellt, die in der Topic-Dokument-Verteilung um die Rechtecke der Dokumente angezeigt werden. Bei kleinerer Topic-Anzahl ist das sinnvoll, bei größerer Zahl kann dies schnell unübersichtlich werden. Teilkorpora können nicht angelegt und dargestellt werden und Dateien können ausschließlich im TXT-Format importiert werden. [20]

### 3.6 Zwischenfazit Analyse-Tools

All diese Analyseangebote bieten zahlreiche Funktionen, um Topic-Modeling-Ergebnisse interaktiv und visuell darzustellen. Jedoch erlaubt es keine der Lösungen, alle vorher herausgearbeiteten Analyseebenen anzusteuern oder interaktiv mit einander zu verknüpfen. Der Import ist fast ausschließlich auf TXT-Formate beschränkt. Das größte Problem besteht darin, dass sie oft auf die eigenen Quellengattungen der Forschungsprojekte angepasst sind und die Möglichkeit zu chunken fehlt. Ein Analyse-Tool für lebensgeschichtliche Interviews muss die Topic-Modeling-Ergebnisse auf der Korpus-, der Archiv-, der Interview- und der Chunk-Ebene darstellen und interaktiv miteinander verknüpfen sowie die Dokument-Ansicht samt Metadaten wie Sprecher\*innenkürzel oder Timecodes anzeigen können. Um diese spezifischen Anforderungen ermöglichen zu können, wird das OHTM-Dashboard als neue Analyselösung präsentiert. [21]

## 4. Das OHTM-Dashboard

Da meine Forschungsarbeit und damit die Entstehung des Tools im Rahmen des von der DFG geförderten Projekts ›Oral-History.Digital‹ entstanden ist, lagen die Interviewtranskripte im Exportformat des Archivs als CSV-Datei vor. Diese Datei enthält in den verschiedenen Spalten die Timecodes, Sprecher\*innenkürzel und Textpassagen. Diese vorgegebene Struktur macht einen einheitlichen Import in die Topic-Modeling-Pipeline sehr einfach und erleichtert das Auslesen der Metainformationen von Timecodes und Sprecher\*innenkürzeln, die für die Berechnung keine Rolle spielen, aber für die Auswertung wichtig sind. Ein Grundproblem von Topic Modeling ist, dass vor der eigentlichen Berechnung die zugrunde liegenden Daten formell aufbereitet und oft individuell abgestimmt werden müssen. Da die Verbindung zwischen Sprecher\*innenkürzeln und Timecodes für die Analyse von Interviews wichtig ist, wäre es viel zu umständlich, die Transkripte in TXT-Dateien umzuwandeln und die Informationen herauszuparsen. [22]

Um den spezifischen Anforderungen lebensgeschichtlicher Interviews zu entsprechen und alle relevanten Ebenen analysieren zu können, wurde ein eigenes interaktives visuelles Analyse-Tool entwickelt. Dafür müssen die Dateien jedoch bereits vor der eigentlichen Topic-Modeling-Berechnung strukturiert und aufeinander beziehbar abgespeichert werden, da dies mit den reinen Mallet-Ergebnissen nicht möglich ist. Deshalb ist das entwickelte Analyse-Tool nicht losgelöst von der dazugehörigen Topic-Modeling-Pipeline nutzbar, da dort die Grundlage für die visuelle Analyse geschaffen werden musste. Die Oral-History-Topic-Modeling-Pipeline (OHTM-Pipeline) ist eine Python-Pipeline, die auf Grundlage von LDA-MALLET (aufbauend auf gemeinsamen Vorarbeiten mit Dennis Möbus) entwickelt wurde.<sup>22</sup> Die Pipeline speichert das Korpus, die Topic-Modeling-Ergebnisse und alle relevanten Einstellungen in einer einzigen Datei, dem OHTM-File, das eine eigens entwickelte JSON-Datenstruktur enthält. Auf Grundlage dieser Datenstruktur wurde das auf dem Python-Paket Plotly basierende Dashboard<sup>23</sup> entwickelt, das mit dem Python-Skript des OHTM-Dashboards geöffnet wird. Um das Tool auch für Forschende anderer Disziplinen nutzbar zu machen, kann die Pipeline mit einfachen TXT-Dateien genutzt werden. Sollten die Daten spezifische Eigenheiten mitbringen, kann der Import innerhalb der Pipeline individuell angepasst werden. [23]

Der Vorteil und das Alleinstellungsmerkmal des OHTM-Dashboards sind die interaktive Verknüpfung aller Analyseebenen untereinander sowie die stetige Bezugnahme zu den Interviews des Korpus, um die wechselseitige Exploration (oder auch Evaluation) zu ermöglichen. Im Folgenden wird das Dashboard samt seinen einzelnen Funktionen vorgestellt, beginnend bei der obersten Ebene, der Topic-Korpus-Verteilung, hinunter bis zur Dokumentenansicht. Die Begriffswahl entspricht der Quellengattung dieser Forschungsarbeit. Da die einzelnen Interviews geschunkt wurden, besteht die unterste Ebene, die Dokumentenebene, aus den einzelnen Chunks. Darauf folgt die Interviewebene. Die einzelnen Teil-Korpora entsprechen den Archiven, aus denen die Interviews stammen, und das Korpus umfasst alle Interviews. [24]

### 4.1 Bar-Graph-Ansicht

Die Navigation innerhalb des Dashboards erfolgt über den »Menü«-Button auf der linken Seite, wie Abbildung 10 zeigt. So lassen sich die verschiedenen Analyseebenen ansteuern. [25]

<sup>22</sup> Vgl. Möbus 2020, S. 162–180; Hodel et al. 2022; Franken / Möbus 2024; Bayerschmidt 2026.

<sup>23</sup> Vgl. Bayerschmidt 2026b. Das Dashboard ist frei zugänglich und kann online auf der Archivseite des ADG ausprobiert werden: <https://ohtm-dash.oral-history.digital/>.

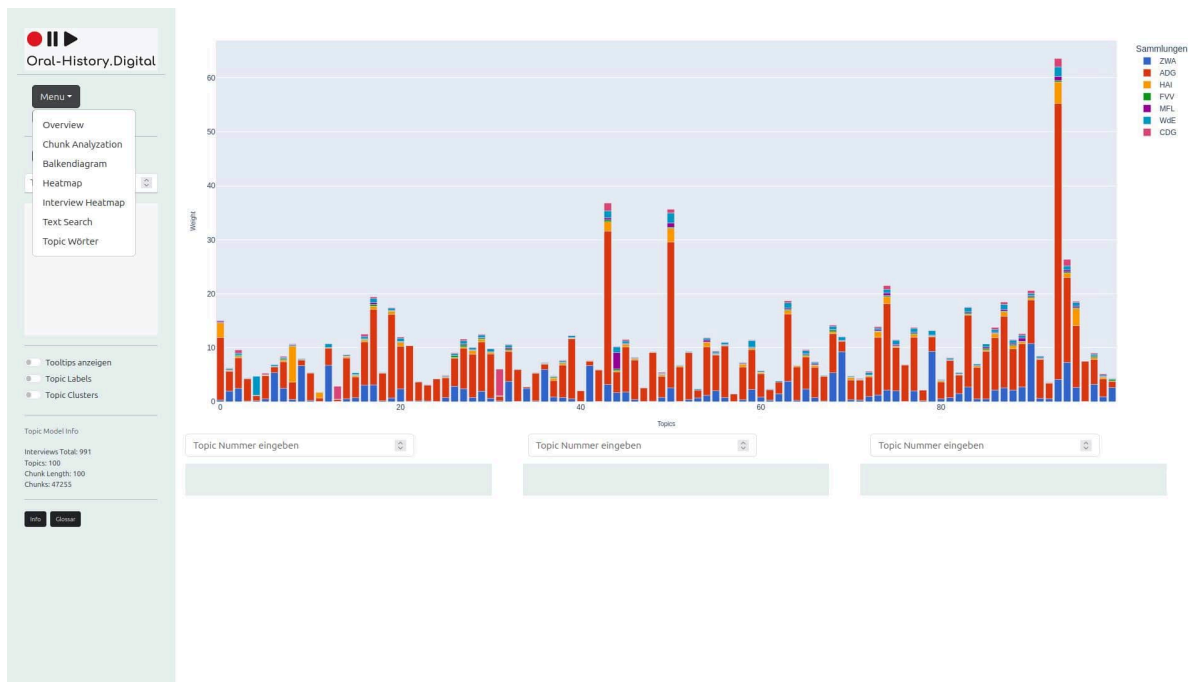


Abb. 10: Navigationsmenü im Menüband. [Screenshot: Philipp Bayerschmidt 2026]

Die »Bar-Graph«-Ansicht (Abbildung 11) zeigt ein Balkendiagramm, welches die Topic-Korpus-Verteilung präsentiert und neben der Gesamtgewichtung (Höhe der Balken) auch die Archive berücksichtigt und abbildet (farbliche Teilbalken). [26]



Abb. 11: Topic-Korpus-Verteilung in der »Bar-Graph«-Ansicht. [Screenshot: Philipp Bayerschmidt 2026]

Die einzelnen Archive können über die Legende angesteuert und separat ein- oder ausgeblendet werden. Unterhalb des Balkendiagramms können in den drei Eingabezeilen Topic-Nummern eingetragen werden, um anschließend die ersten dreißig Wörter des entsprechenden Topics auszugeben. Mit der Maus kann über [27]

die einzelnen Balken gefahren werden, um weitere Informationen zu erhalten. Durch einen Klick auf einen Balken wird das entsprechende Topic auf der linken Seite im Menüband mitsamt der ersten zwanzig Wörter ausgegeben. Diese Funktion ist in jedem Graphen des Dashboards integriert, um die jeweilige Topic-Nummer jederzeit ohne große Umstände auf die jeweiligen Wörter zurückzuführen.

## 4.2 Topic-Wörter

Die Topic-Wort-Verteilung kann auf der Seite »Topic-Wörter« eingesehen werden, indem man die Anzahl der gewünschten Wörter einträgt, woraufhin diese als Liste angezeigt werden. Da die tatsächliche Gewichtung der Wörter eine geringere Rolle spielt als das Zusammenspiel der Wörter, wurde auf die Ausgabe der tatsächlichen Gewichtung verzichtet. Innerhalb des OHTM-Files ist diese Information hinterlegt.

[28]

Topic	Words
0	ddr, berlin, westen, wende, ostern, mauer, eltern, westberlin, prinzip, relativ, lachen, erinnern, spannend, zeit, grenze, kind, passieren, ostberlin, erleben, interessant, merken, 89, freund, mitbekommen, familie, super, stadt, kennen, rüber, dresden, west, berliner, kontakt, stasi, politisch, richtung, spät, westdeutschland, gefühl, drüben, richtig, leipzig, system, land, wiedervereinigung, normal, besuchen, angucken, wahnsinnig, echt,
1	arbeit, maschine, fabrik, meister, werk, betrieb, arbeiter, bauen, herstellen, schwer, werkstatt, stunde, schicht, verschieden, zeit, lernen, schlosser, ingenieur, hand, halle, abteilung, vorarbeiter, produktion, tag, reparieren, einstellen, krieg, einsetzen, chef, technisch, arbeitsplatz, werkzeuge, akkord, drehen, firma, kaputt, metall, verdienen, ordnung, anlage, produzieren, material, reparatur, dreher, spät, betreiben, schmied, kennen, büro, beschäftigen,
2	meter, bauen, stein, holz, arbeit, schwer, müssen, hand, wasser, erde, richtig, loch, berg, dick, nennen, graben, kiste, fallen, stelle, zeit, eisen, breit, sand, tag, draht, fest, boden, sammeln, wagen, bretter, zement, bau, maschine, hart, zentimeter, 50, luft, löcher, spät, brechen, kopf, kilo, verlegen, straße, baum, winter, hacken, material, passieren, ring,
3	krupp, essen, abteilung, chef, firma, herr, betrieb, büro, krieg, thyssen, spät, lachen, kollege, direktor, gehören, kennen, beschäftigen, sekretärin, zeit, kruppsch, arbeit, angestellter, duisburg, heimat, erde, mülheim, dr, personalabteilung, nennen, arbeiter, einstellen, entlassen, hütte, übernehmen, kruppiener, verwaltung, arbeitsamt, stelle, essener, sozial, arbeitsplatz, hamborn, borbeck, schlecht, abteilungsleiter, verhältnis, damen, werk, doktor, gehalt, altendorf,
4	hamburg, italienisch, italiener, italien, deutsch, hamburger, firma, arbeiter, bahn, deutschland, mitarbeiter, spät, herr, altona, ausländer, konsulat, harburg, ausländisch, lübeck, sogenannten, gewohnen, barmbek, eppendorf, arbeit, bahnhof, kühn, organisieren, ipuf, weltkrieg, tätigkeit, wilhelmsburg, arbeitsamt, mussolini, zweiter, kennen, türke, gastarbeiter, siebziger, sizilien, hafen, bergedorf, käthe, anfang, bundesbahn, kj, sechziger, ochsenszoll, mailand, behörde, wandsbek,
5	gericht, akte, prozess, herr, richter, urteil, verurteilen, anwalt, stasi, rechtsanwalt, politisch, ddr, gesetz, verfahren, staatsanwalt, cottbus, berlin, ehemaling, grund, bericht, antrag, beantragen, verhandlung, aussage, paragraph, justiz, geschichte, ablehnen, ministerium, jurist, zeuge, staatsicherheit, deutsch, landgericht, herm, natke, vertreten, hälftling, haft, erklären, stimmen, bundesrepublik, strafe, aussagen, prozesse, entschädigung, begründung, behörde, rehabilitierung, melden,
6	jüdisch, jude, israel, erinnern, hebräisch, religiös, synagoge, familie, lernen, kind, stadt, gemeinde, jiddisch, nennen, rabbiner, ort, vater, holocaust, zeit, schauen, ungarisch, orthodox, kennen, geschichte, bruder, schule, synagog, mensch, ungar, rabbi, feiertag, schabbat, palästina, mädchen, spät, beten, kippur, jerusalem, gebet, religion, befinden, pessach, sprache, hören, zuhause, polen, niederlande, helfen, problem, art,
7	nähen, anziehen, schön, schwer, kleidung, hose, schuh, stoff, kleid, rot, mädchen, haar, blau, stricken, kleider, haare, kaufen, anzug, mantel, stiefel, hand, mütter, jacke, hemd, uniform, rock, zeit, bluse, grün, pullover, farbe, hosen, winter, strümpf, tasche, grau, dick, braun, schneider, arm, kopf, erinnern, klamotte, leder, schneiden, lernen, mütze, stopfen, nähmaschine, kittel,
8	kind, eltern, kindergarten, schule, schön, gruppe, wende, krippe, einrichtung, lernen, erzieher, erziehung, ddr, mutti, erzieherin, ddrzeit, gerne, dresden, zeit, kindheit, erinnern, ausbildung, , familie, leiterin, mütter, erleben, hort, freiheit, lehrer, erziehen, betreuen, arbeit, stunde, kindergarte, pionier, klasse, narayen, kindergärt, normal, kinderkruppe, gehören, abholen, spaß, kindergärtnerin, tag, ferienlager,

Abb. 12: Topic-Wörter-Ansicht. [Screenshot: Philipp Bayerschmidt 2026]

## 4.3 Heatmap

Die »Heatmap« (Abbildung 13) bietet einen Überblick über die Topic-Interview-Verteilung, die summierte durchschnittliche Topic-Gewichtung pro Interview. Da die Interviews zuvor gechunkt wurden, ermöglicht diese Übersicht einen Einblick in die Topic-Verteilung pro Interview und zeigt, welche Topics im gesamten Interview am höchsten gewichtet sind.

[29]

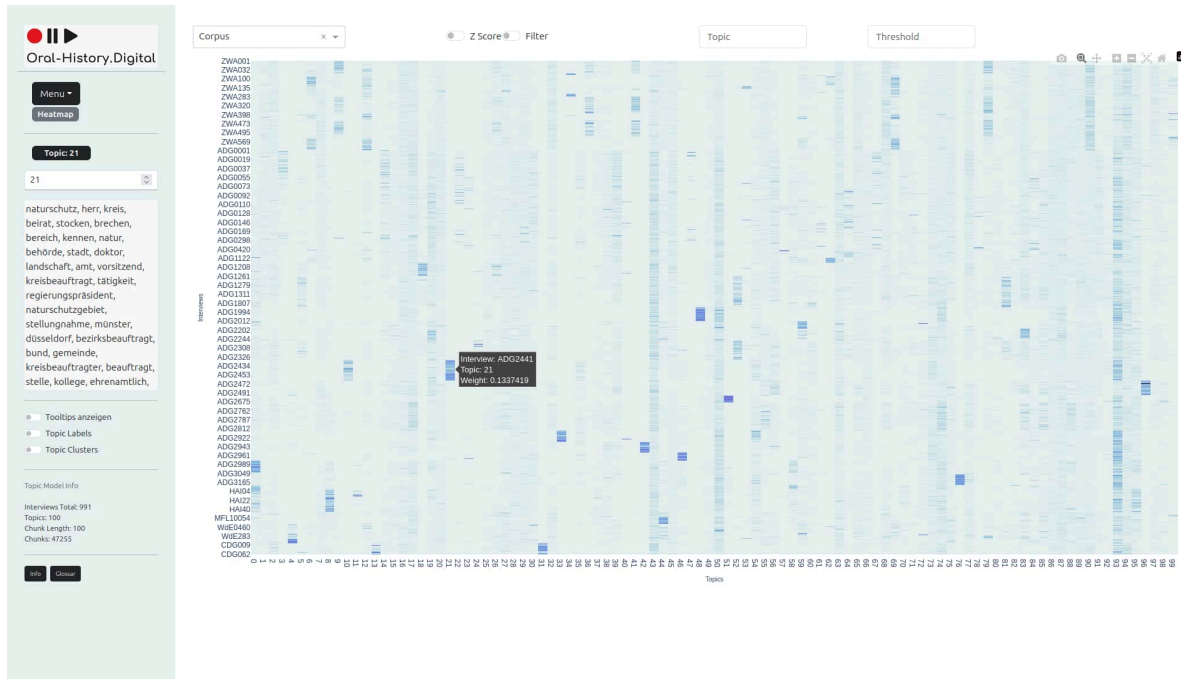


Abb. 13: Topic-Interview-Verteilung in der Heatmap. [Screenshot: Philipp Bayerschmidt 2026]

Über die Auswahl eines Topics in der Grafik durch einen Klick können die entsprechenden Wörter im Menüband angezeigt werden. Die Heatmap kann bequem über das Ziehen eines Feldes mit der Maus gezoomt werden. Über das Dropdownmenü »Corpus« können die einzelnen Archive ausgewählt und die Heatmap dementsprechend gefiltert werden. Eine weitere Filteroption, auswählbar über den Schalter »Filter«, bietet die Möglichkeit, die Heatmap nach einem bestimmten Topic samt *Threshold* filtern zu lassen. [30]

Da sich die Heatmap in der Darstellung an maximalen und minimalen Werten orientiert und die Gewichtung aller Werte in entsprechenden Relationen dargestellt wird, können bei starken Extremen die einzelnen Werte verblassen und undeutlich erscheinen. Mit dem Schalter »Z-Score« wird eine Standardabweichung der Werte berechnet, um die Spitzen zu glätten und die Darstellung deutlicher zu machen. Dadurch wird die Lesbarkeit und Aussagekraft der Heatmap für das gesamte Korpus erhöht und die Identifikation von allgemeinen Outliern erleichtert. [31]

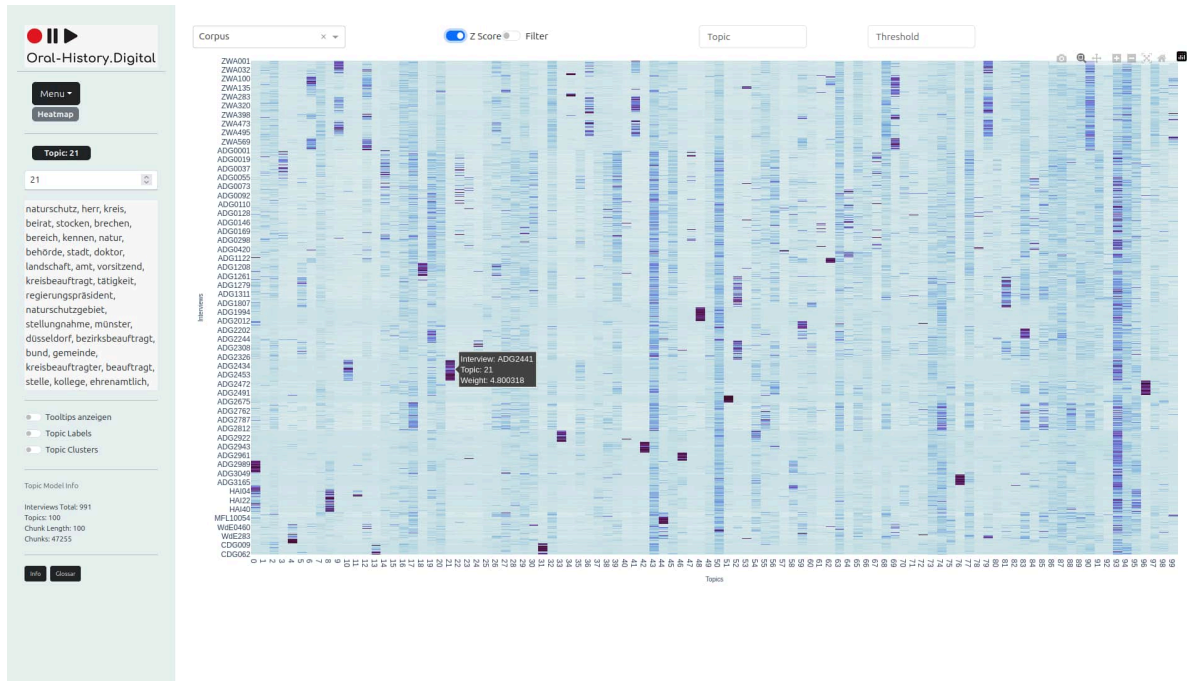


Abb. 14: Z-Score-bereinigte Heatmap. [Screenshot: Philipp Bayerschmidt 2026]

### 4.4 Interview-Heatmap

Eine Ebene weiter gelangen wir zur »Interview-Heatmap« (Abbildung 15), die die Topic-Chunk-Verteilung der jeweiligen Interviews anzeigt.

[32]



Abb. 15: Interview-Heatmap, Interviewausschnitt adg0001. [Screenshot: Philipp Bayerschmidt 2026]

Dennis Möbus entwickelte im Rahmen von OH.D die Interview-Heatmap, in der die Achsen getauscht und die Chunks in chronologischer Reihenfolge auf der x-Achse abgebildet werden, um so den Topic-Verlauf im Interviewverlauf visualisieren zu können.<sup>24</sup> Durch die Integration in das Dashboard können mit einem Klick auf das entsprechende Chunk die zum ausgewählten Topic gehörenden Wörter im Menüband angezeigt und die dazugehörige Transkriptpassage in unverarbeiteter Form samt zugehörigen Sprecher\*innenkürzeln unterhalb der Heatmap ausgegeben werden. Wenn der Schalter »Marker« aktiviert ist, wird der ausgewählte Chunk farblich markiert. Zusätzlich werden der Timecode des entsprechenden Chunks sowie, wenn rechtlich und technisch möglich, ein Direktlink zum entsprechenden Audiofile der Textstelle im OH.D-Archiv angezeigt.

Über die beiden Pfeiltasten kann im Interviewtranskript geblättert und der vorherige und nachfolgende Chunk samt Text angezeigt werden, der Marker bewegt sich mit. Eine Angleichung der Ergebnisse über die Berechnung des Z-Scores ist erneut möglich, allerdings fällt die Verbesserung deutlich geringer aus als bei der Korpus-Heatmap.

### 4.5 Overview

Das Kernstück des OHTM-Dashboards findet sich in der »Overview«-Seite (Abbildung 16), die die verschiedenen Analyseebenen gleichzeitig zugänglich macht und eine kombinierte, interaktive Nutzung ermöglicht. Dadurch wird ein scalable reading in der Kombination aus close und distant reading möglich.

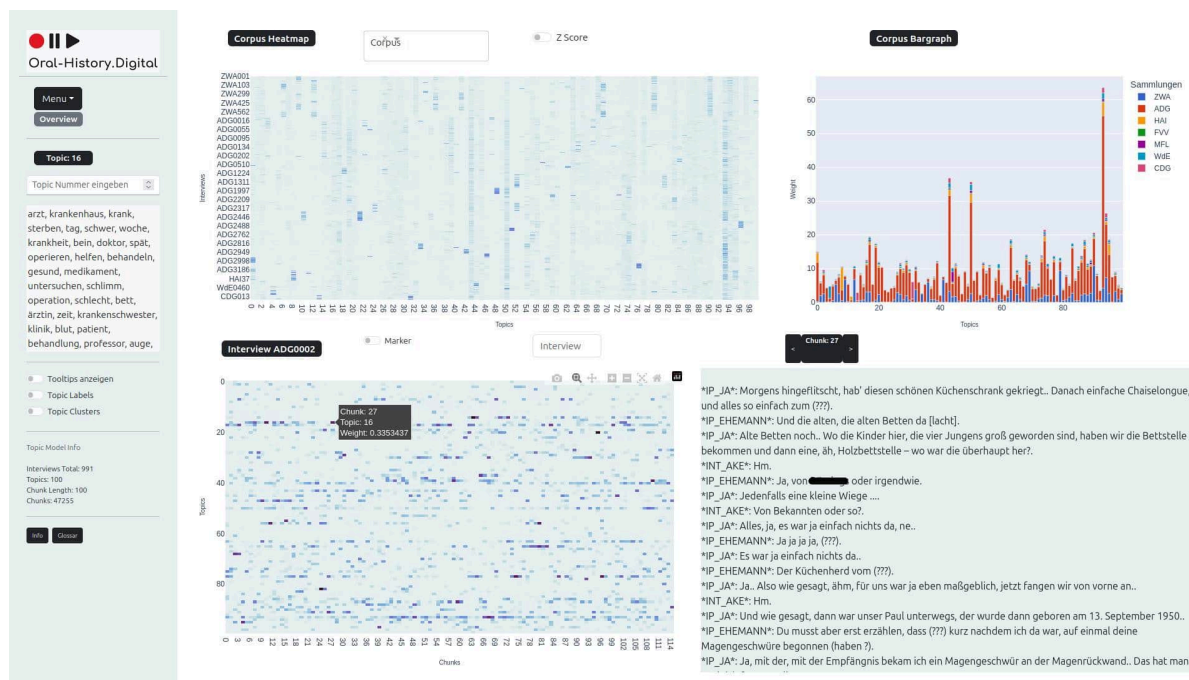


Abb. 16: Analyseebenen in der Overview-Ansicht, Interviewausschnitt adg0002. [Screenshot: Philipp Bayerschmidt 2026]

In jedem der Graphen ist es möglich durch einen Klick auf das Topic die entsprechenden Wörter im Menüband anzeigen zu lassen. Wenn man auf ein Feld in der Korpus-Heatmap klickt, wird das entsprechende Interview in der Interview-Heatmap angezeigt. Durch die Auswahl eines Chunks dieses Interviews kann der unverarbeitete Text ausgegeben werden, der sich über die Pfeiltasten navigieren lässt.

<sup>24</sup> Vgl. Möbus 2025.

Diese Kombination ermöglicht es auf explorative Weise die Ergebnisse entsprechend der Forschungsfrage zu analysieren. Wurde ein passendes Topic gefunden, kann ausgelesen werden, wie das Topic im gesamten Korpus oder in einzelnen Archiven vertreten ist, welche durchschnittliche Gewichtung es in den einzelnen Interviews besitzt und wie die tatsächliche Verteilung innerhalb eines Chunks aussieht. Die Chunk-Ansicht kann Informationen zur Korrelation von Topics bei aufeinanderfolgenden Chunks liefern und so Erkenntnisse über vorausgegangene oder nachfolgende Themen liefern. Verschiedene Interviews und die entsprechenden Textstellen können miteinander verglichen werden, wobei die jeweiligen Topics stets auf die zugehörigen Wörter zurückgeführt werden können. [37]

Über die drei Optionsschalter im Menüband können zusätzliche Optionen für alle Seiten des Dashboards aktiviert werden. Mit »Tooltips anzeigen« werden beim Hovern über den verschiedenen Feldern Texte zur Funktionalität eingeblendet. [38]

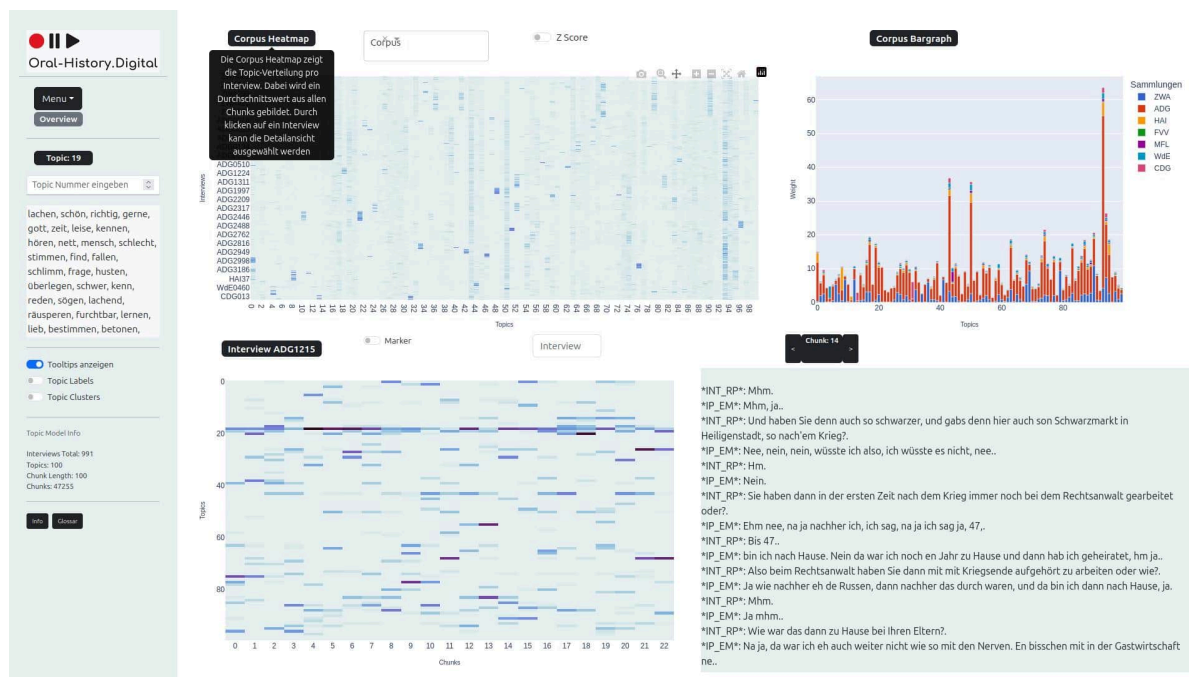


Abb. 17: Tooltip-Anzeige im Dashboard, Interviewausschnitt adg1215. [Screenshot: Philipp Bayerschmidt 2026]

In der OHTM-Pipeline können die Topics nach dem Trainieren eines Modells von Forschenden mit Schlagwörtern bzw. Labels versehen werden. Durch die Aktivierung von »Topic Labels« werden diese Schlagwörter innerhalb der Ausgabe im Dashboard zusätzlich zu den Topic-Nummern angezeigt. Da die Labels eine Interpretation darstellen, ist es wichtig, dass die Topics stets auf die Topic-Wörter zurückgeführt werden können. [39]

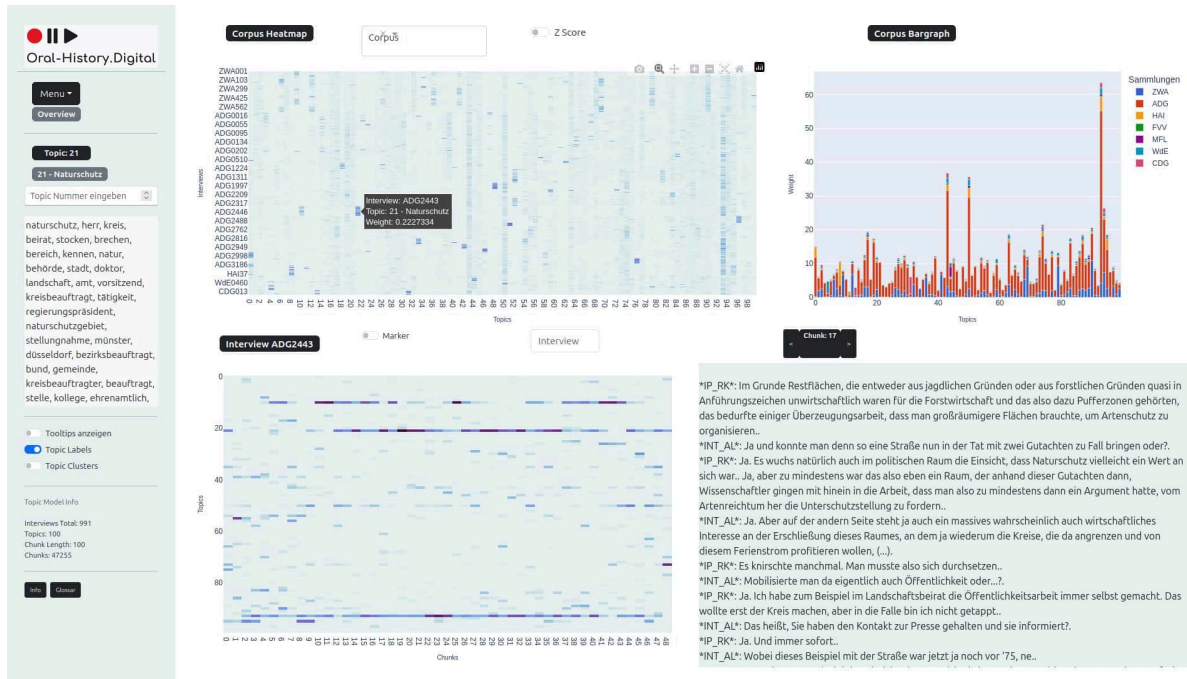


Abb. 18: Label-Anzeige für die einzelnen Topics, Interviewausschnitt adg2443. [Screenshot: Philipp Bayerschmidt 2026]

Bei einer großen Topic-Anzahl können mehrere Topics ähnliche Themen abdecken, beispielsweise mehrere Topics zum Themenkreis »Familie«. Diese können in der OHTM-Pipeline händisch zu selbstbenannten Clustern zusammengefügt werden, sodass beispielsweise anstatt der hier im Beispiel vorhandenen 100 Topics nur 25 Topic-Cluster angezeigt werden.

[40]

Unterhalb dieser Optionen werden einige generelle Informationen zum geladenen Topic Model angezeigt. Im Impressum und Glossar können über eine Textdatei Informationen für Nutzende hinterlegt werden, die durch einen Klick geöffnet werden kann.

[41]



Abb. 19: Cluster-Anzeige, Interviewausschnitt adg2467. [Screenshot: Philipp Bayerschmidt 2026]

### 4.6 Chunk-Suche und Chunk-Analyse

Zusätzlich bietet das Dashboard zwei weitere Analyseebenen, die es ermöglichen, das Forschungskorpus gezielt nach passenden Textstellen zu durchsuchen. [42]

Über die »Chunk-Suche« können entweder die Chunks des gesamten Korpus (»Korpus Search«) oder eines einzelnen Interviews (»Interview Search«), einstellbar über das Dropdown-Menü, nach einem Topic samt Schwellenwert gesucht werden. [43]

weight	Interview	Chunk Nr	Chunk	Top 5 Topics
0.5990997305708491	MFL	23	Wir verstehen uns, das machen viele Russen so. Und ähm... *BV* Ich mache nicht so. Ich spreche auf Russisch, dann spreche ich auf Russisch ohne deutsche Wörter. Ich spreche auf Deutsch, ja nicht perfekt, nicht richtig, aber ich spreche auf Deutsch. Ja, so. *CV* Also für mich manche Sachen kann ich nicht auf Russisch erklären. Ich kann auf Deutsch, aber auf Russisch nicht. Also doch kann ich schon aber schwierig. *BV* Ach, nächste Frage, bitte. Haben Sie... *CV*... *Int*... *DP*... *CV* Das ist mein Handy... *DP* Ja ihr habt ja jetzt viel erzählt, wie es am Anfang war. Wie ist es mittlerweile für euch, wie ist jetzt euer Eindruck von Deutschland... *CV* Perfekt. Das ist bestes, was mit uns passieren könnte. Wir sind echt sehr sehr zufrieden... *BV* hm... *CV* Wir mögen dieses Land. Wir wollen nicht zurück oder nach irgendwelches noch Land. Nein. Wir haben uns hier gefunden, also ich, für mich kann ich sagen ich bin daheim. Dahoam, oder wie sagt man des, ja... *DP* Dahoam... *CV* Dahoam, ja, ja. Genau... *BV* Ja, mir gefällt auch. Ich habe über Deutschland geträumt. Ich wohne da... *CV* Ne... *DP* Und is, is es in etwa so wie Du es dir vorgestellt hast, oder... *BV* Nein, ah, ich habe mir über deutsche Leute gedacht, zum Beispiel, gibt es einige intelligente, zum Beispiel freundlich wie in Kreis in... zum Beispiel gewohnt, aber andere... *CV* Das sind nicht, nicht Deutsche, dass sind Leute wir sagen, dass kann man nicht an der Nationalität ähm... *BV* Nu Leute, ja. Gibt es freundlich, gibt es unfreundlich und ja, aber ich habe gedacht, ein bisschen mehr freundlich müssen Deutschen sein... *CV* Der meint vielleicht, wir spüren, dass, an uns sozusagen, dass wir Ausländer sind und das kommt ein bisschen von anderen Leuten. Leute die nicht Ausländer mögen es gibt ja in jede Land, in jedem Land gibts solche Leute. Und da gibts ja auch die nicht zum Beispiel Russen mögen, es gibt auch. Und man spürt das. Es gibt sowas auch... *BV* Nun, nein, zum Beispiel ich verstehe einige Wörter nicht und das war schwer. Ich habe Führerschein gemacht, bestanden, aber Lehrerin war Oh du verstehst nicht, du verstehst nicht. Ich habe gesagt ich lerne. Ich wohne nur in Deutschland ein Jahr. Was, muss ich perfekt sprechen, oder was. Ja, ich werde weiter lernen, ich freundlich und so weiter. Ich zum Beispiel, ich helfe immer. Wenn man braucht zum Beispiel braucht im Fitnessstudio Nachbar ich habe alles getragt. Wir sind nach Bar ähm zum Bank zum Beispiel gefahren und wir haben alles geholfen, aber andere zum Beispiel Hilfe Andere Oh Sie sprechen nicht perfekt. Tschüss... *CV* Ich glaube, man wird so behandelt, also wenn man hört wie wir auf Deutsch reden, ja. Und dann bildet man so ein Bild im Kopf, ja. Also, der redet nicht so gut Deutsch, vielleicht der ist dumm oder so oder keine Ahnung. Und so ein Eindruck war von uns vielleicht auf am Anfang. Und das ist ja vielleicht bis jetzt noch so, aber das heißt nicht, dass wir dumme Leute oder so, also wir haben ja auch die Schule besucht, wir können auch Schreiben, Lesen und wir können auch, wir sind auch in andere Richtungen gut, vielleicht wie andere Leute oder so, aber das ist das Problem. Man, man... *BV* hm... *CV* Man macht den ersten Eindruck von dem wann man sieht und wenn man was hört von dem Mensch. Und das ist natürlich, glaube ich, war der Fall, ja. Weil wir haben natürlich im Jobcenter oder so, wir haben, wir könnten nicht so gut reden, aber irgendwie, ich hab glaube ich, ich kann sagen, dass ich Deutsch gelernt habe, so, ich hab nur einfach gesagt, was ich denke. Es war mir egal, wie ich den Satz baue. Bei mir war Nomen am Ende, Verb am Anfang, das war mir egal. <a href="https://portal.oral-history.digital/mfl/de/interviews/mfl-">https://portal.oral-history.digital/mfl/de/interviews/mfl-</a>	44: 0.5990997305708491   94: 0.21392577268082863   0: 0.04873578232889533   87: 0.02992986020180754   50: 0.020996276856984443
0.5933692597528868	MFL	26	Also bei uns ging das ratz-fatz muss ich sagen. Die ganze Sache. Weil alles geklappt hat. Wir konnten Friedland alle quasi Deutsch. Und ich habe meine Kinder vor unserer Reise für ein dreiviertel Jahr in die gr. Deutsch-Sprache-Kursus geschickt. Dass sie wenigstens etwas Ahnung haben, ja. Na und wir beide mit der Frau, wir haben ja nur Deutsch gesprochen. Und wir, bei uns war fest, wenn wir die Grenze überschreiten, dann reden wir nicht ein einziges polnisches Wort miteinander und das ist bis heute so. *Int* Warum... *PU* Ja, weil wir eben wollten, so lange nur Deutsch reden bis die polnische Sprache eine Auslandsprache für uns ist. Also wenn jemand kommt jetzt auf, aus Polen dann, ich habe Schwierigkeiten mich zu äußern, alles, wenn die, die Sprache, ich will nicht als Pole existieren in Deutschland. Das ist für mich die, die Grundlage alle, alle, die jetzt, sage mal schließlich bekennen zu. Ich hatte die Frage bei meinem 18ten Geburtstag musste ich bei den Behörden und was sehr gefürchtet war von allen erklären, was für eine Nationalität ich bin, ja. Und viele hatten Angst. Also da hat, hat man die Staatsangehörigkeit angeben, natürlich staatsangehörig war ich polnisch, weil ich dort gelebt habe. Aber die zweite Frage war Nationalität. Und da haben viele, die sich als Deutsche	44: 0.5933692597528868   74: 0.10623190637797401   43: 0.0784961979980361   93: 0.050625656043916625   63: 0.03911055233983686

Abb. 20: Textsuche in den Chunks, Interviews des Museums Friedland. [Screenshot: Philipp Bayerschmidt 2026]

Anschließend werden alle Ergebnisse in einer Tabelle mitsamt des Topic-Gewichts, der Interview-ID, der Chunk-Nummer, dem unverarbeiteten Text und den ersten fünf Topics des Chunks ausgegeben. Eine weitere Analysemöglichkeit erweitert die Chunk-Suche, indem nicht nur der Text ausgegeben wird, sondern ähnlich wie in der Overview alle relevanten Ebenen in die Ergebnispräsentation eingebunden werden, die »Chunk-Analyzation«-Seite.

[44]

**ADG0021-Chunk: 51**

\*IP\_HH\*: Wenn sie wieder äh hinkamen, dann waren sie Mitinhaber einer Firma... Und deshalb glaub ich nicht, dass die so wie der Pole, hierherkamen und sich ... (???) da einlebten... Das war ja gar kein Fremdarbeiter im Sinne, der ist umgezogen, in meinen Augen...

\*INT\_UH\*: [lacht] Ja!

\*IP\_HH\*: Der ist umgezogen und hat dann hier gearbeitet! Ist dann, hat dann die deutsche Staatsangehörigkeit gekriegt und so, ist dann richtig hier eingee... hat sich hier eingelebt...

\*INT\_UH\*: Wenn Sie aber an die Kinder der Gastarbeiter heute denken, ja, die Italiener, die wir so an der Schule haben, die leben ... die sind hier geboren, die fühlen sich als Deutsche!

\*IP\_HH\*: Ja! Ja, ich würde sagen, (???) im (???) hier, da sind so ein paar Türkenkinder und so drin, die machen ein Geschrei und ich hab mal miterlebt, wie eine Frau sagt „Nun hört doch mal endlich auf!“, [mimt Geschrei], „Wawawaaa!“, und da sagt ein Mann „Jetzt halte endlich deine Schnauze, sonst kriegste ein paar drauf!“ [mit gepresster, lauter Stimme] „Dann kommt mein Vater, und sein (???) und, und dann konnten sie auf einmal Deutsch! Weiß nicht ... vielleicht haben wir das früher nicht so mitgekriegt mit den Polen oder so, vielleicht waren die auch nicht in so großer Zahl da wie hier ...“

\*INT\_UH\*: Ja endlichherweise. Mhm

44: 0.3072 - 44  
19: 0.1683 - 19  
22: 0.1303 - 22  
30: 0.0935 - 30  
17: 0.0381 - 17  
63: 0.038 - 63  
59: 0.0376 - 59  
24: 0.0373 - 24  
74: 0.0289 - 74  
66: 0.0283 - 66  
50: 0.0202 - 50  
4: 0.0187 - 4  
20: 0.0099 - 20  
10: 0.0095 - 10  
93: 0.0027 - 93  
47: 0.0019 - 47

Topic 44 (116)  
Topic 50 (34)  
Topic 93 (28)  
Topic 87 (22)  
Topic 43 (20)  
Topic 17 (19)  
Topic 89 (19)  
Topic 74 (16)  
Topic 95 (15)  
Topic 54 (15)  
Topic 75 (14)  
Topic 94 (13)  
Topic 88 (13)  
Topic 32 (13)  
Topic 63 (12)  
Topic 19 (12)

Abb. 21: Chunk-Analyzation mit Chunk-Suche und weiteren Analyseebenen, Interviewausschnitt adg0021. [Screenshot: Philipp Bayerschmidt 2026]

Das Korpus kann nach einem bestimmten Topic samt Schwellenwert gefiltert werden. Das Balkendiagramm zeigt die Topic-Verteilung der Ergebnisse, die Heatmap die Topic-Chunk-Verteilung an. Zusätzlich werden die Anzahl der gefundenen Chunks und die Anzahl der jeweiligen Interviews ausgegeben. Wie in der Interview-Heatmap kann durch die Auswahl eines Chunks der unverarbeitete Text angezeigt werden. Daneben finden sich zum einen die gesamte Topic-Verteilung des einzelnen Chunks sowie die Summe aller Top-5 Topics in den Ergebnissen. Wenn man den »Correlations«-Schalter aktiviert, kann die Suche noch um ein zweites Topic samt Parameter als zusätzliche Filterfunktion ausgewählt werden. [45]

## 4.7 Nicht berücksichtigte Funktionalitäten

Die Entwicklung des Dashboards hatte die Kombination von distant und close reading im Blick, um das Wechselspiel von scalable reading zu ermöglichen. Dabei wurden einige Funktionalitäten nicht umgesetzt, weil sie für dieses Ziel nicht zwingend notwendig erschienen. Dazu zählt die aus dem LDAvis oder Topic Cloud bekannte Topic-Wort-Verteilung, die die Verbindungen der Topics und gemeinsamer Wörter untereinander darstellt. Des Weiteren können innerhalb der Dokument-Ansicht die Wörter nicht entsprechend ihrer zugehörigen Topics oder Stopwords farblich markiert werden, wie es beispielsweise im TwiC möglich ist. Ebenso wurde darauf verzichtet, die Gewichtung der Wörter in der Topic-Wort-Ansicht auszugeben. [46]

## 5. Topic ›Beheimatung‹ – ein Use Case

In meinem Promotionsprojekt *Topic ›Beheimatung‹* analysiere ich lebensgeschichtliche Interviews der Oral History hinsichtlich enthaltener Erzählungen über die Themen ›Heimat‹ und ›Beheimatungsprozesse‹<sup>25</sup>. Dabei soll untersucht werden, wie Menschen, die nach Deutschland gekommen und hier geblieben sind, innerhalb ihrer Lebensgeschichten über diese Themen erzählen. Der Fokus liegt weniger auf der Einordnung in die historische Entwicklung des Begriffs ›Heimat‹<sup>26</sup>, sondern vielmehr auf den individuellen Erzählungen und den Aspekten des persönlichen Verständnisses sowie der eigenen Gestaltung von Heimat.<sup>27</sup> Welche Erfahrungen haben die Menschen gemacht und wie lassen sich die verschiedenen Erzählungen dieser Migrant\*innen auch mit Bezug auf die jeweiligen historischen Rahmenbedingungen vergleichen? Mithilfe von Topic Modeling und dem OHTM-Dashboard wird das zuvor vorgestellte Forschungskorpus aus 991 Interviews hinsichtlich dieses Forschungsinteresses analysiert. Dabei spielte das Dashboard nicht nur nach der finalen Berechnung eine Rolle, sondern auch schon zuvor. Zur Bestimmung der verschiedenen Topic-Modeling-Parameter (Topic-Anzahl, Chunk-Länge, Iterationen) ist es wichtig, die Ergebnisse zu analysieren und durch das Vergleichen verschiedener Modelle die idealen Werte zu finden. Über das Dashboard konnten visuelle Vergleiche zur Verteilung der Topics über das Korpus durchgeführt werden. So erlaubten die verschiedenen Visualisierungen (z. B. Balkendiagramm und Heatmap) in Kombination mit einem close reading bestimmter Chunks und der Überprüfung der zugeordneten Topics die Herausarbeitung wichtiger Erkenntnisse. Somit half das Dashboard schon während der Berechnung des Topic Models.<sup>28</sup> [47]

Über die Topic-Wörter-Ansicht konnte im OH.D-Modell das Topic 44 ausfindig gemacht werden, welches aufgrund seiner Wörter inhaltlich passend zum Thema ›Beheimatung‹ ist. Ziel war es dann, über die Topic-Dokument-Verteilung die Interview-Chunks zu finden, in denen dieses Topic dominant ist und damit die Textpassagen, in denen über Heimat und Beheimatung erzählt wird. [48]

<sup>25</sup> Vgl. Mitzscherlich 2019, S. 183–196.

<sup>26</sup> Vgl. Kück 2022; Oesterhelt 2021; Scharnowski 2019.

<sup>27</sup> Vgl. Jäger 2017, S. 12; Demantowsky 2021, S. 187; Mitzscherlich 2019, S. 184, 187–188.

<sup>28</sup> Vgl. Bayerschmidt / Möbus 2026.

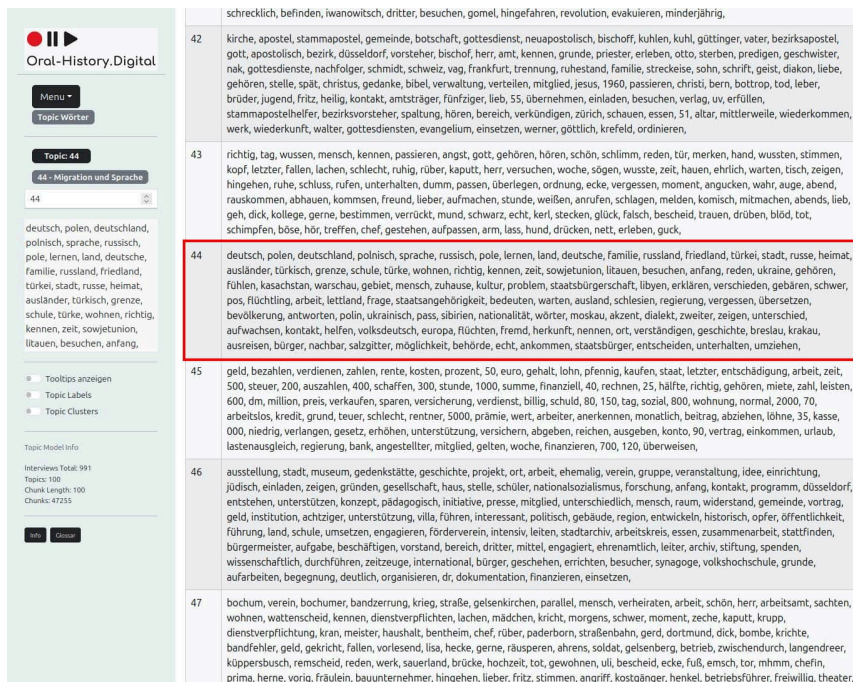


Abb. 22: Topic-Wörter-Ansicht mit markiertem Topic 44. [Screenshot: Philipp Bayerschmidt 2026]

Topic-Nummer	Ersten 20 Wörter des Topics
44	deutsch, polen, deutschland, polnisch, sprache, russisch, pole, lernen, land, deutsche, familie, russland, friedland, türkei, stadt, russe, heimat, ausländer, türkisch, grenze,

Tab. 3: Topic-Wort-Verteilung des Topics 44.

Das Balkendiagramm (Abbildung 23) zeigt über die farbliche Unterteilung sehr deutlich, dass das Topic 44 in mehreren Teilkorpora vorkommt und so eine sammlungsübergreifende Suche möglich ist: ZWA, ADG, MFL, WDE. [49]

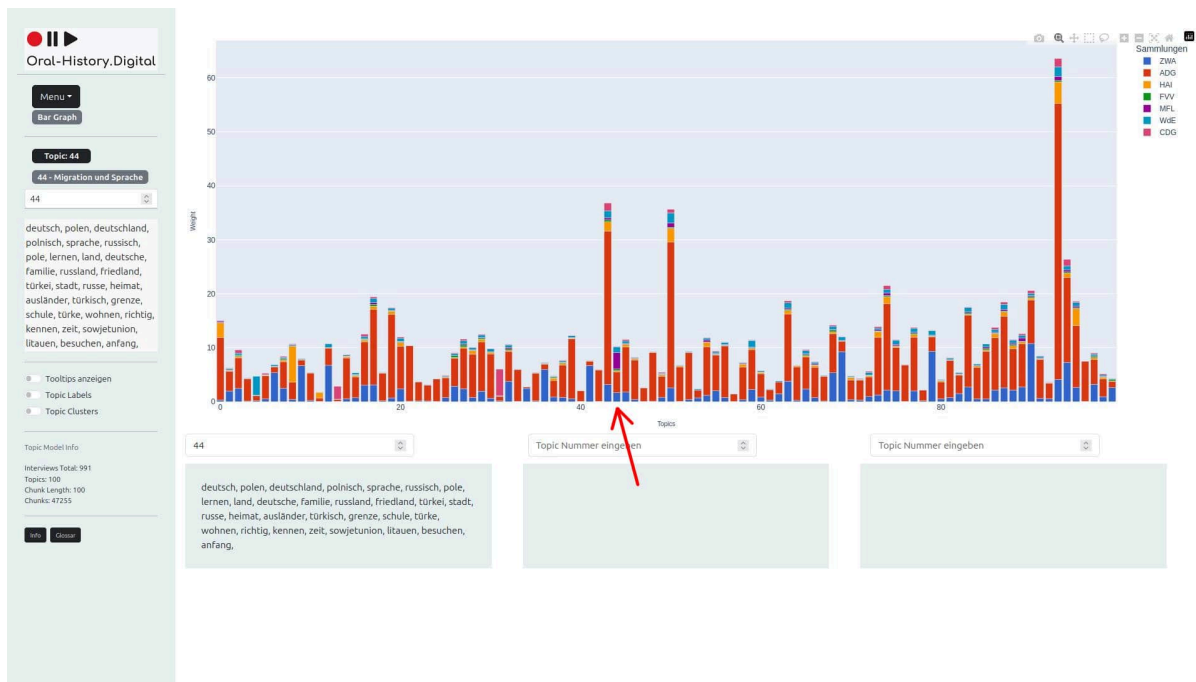


Abb. 23: Balkendiagramm mit markiertem Topic 44. [Screenshot: Philipp Bayerschmidt 2026]

Die Heatmap zeigt, dass es einige Interviews innerhalb des Korpus gibt, in denen Topic 44 im Durchschnitt markant präsent ist. Das Cluster im unteren Bereich der Heatmap zeigt, dass das Topic in den Interviews einer Sammlung besonders stark vertreten ist. Diese Informationen sind ein erster Hinweis darauf, welche Interviews eventuell mehrere Erzählpassagen zum gesuchten Thema enthalten, das sich in der durchschnittlichen Gewichtung wiederfindet.

[50]

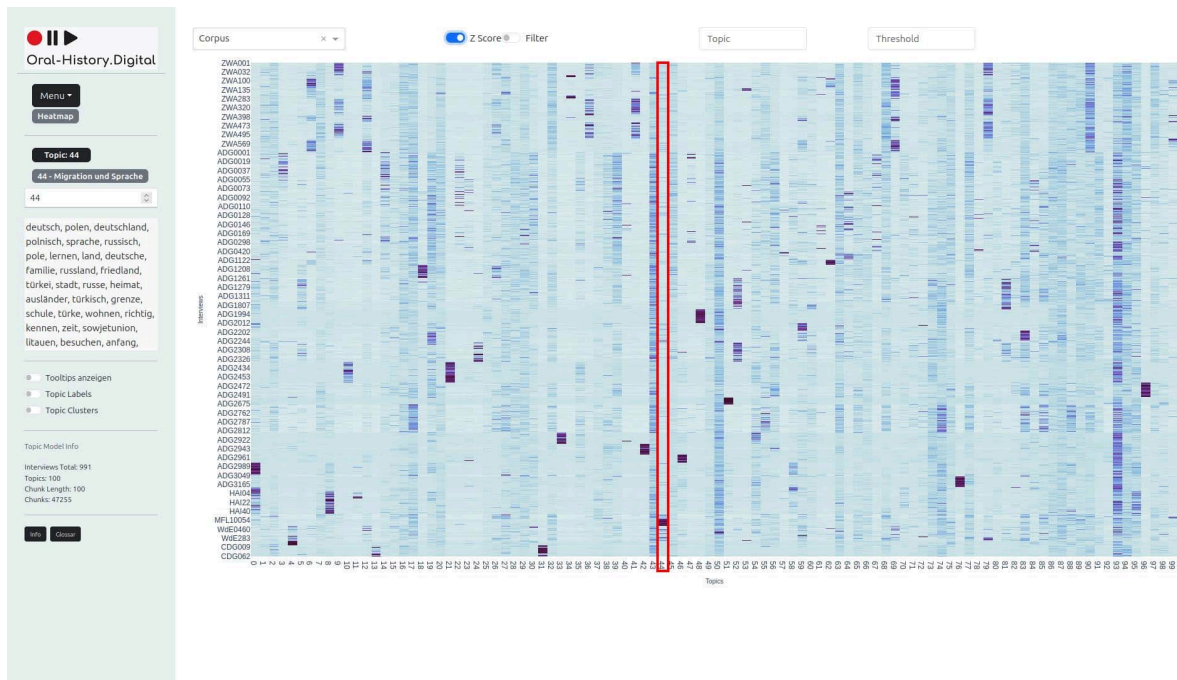


Abb. 24: Heatmap mit markiertem Topic 44. [Screenshot: Philipp Bayerschmidt 2026]

Nachdem über die Textsuche die Chunks gefiltert wurden, in denen Topic 44 den höchsten Wert besitzt, zeigt das höchste Ergebnis eine interessante Textpassage. [51]

The screenshot shows the Oral-History.Digital search interface. On the left, there is a sidebar with a search bar, filters for 'Topic 44' and '44 - Migration und Sprache', and a list of languages. The main area displays search results in a table with columns for 'weight', 'Interview', 'Chunk Nr', 'Chunk', and 'Top 5 Topics'. The top result (Chunk 23) shows a text passage with several words underlined in red, indicating a match with Topic 44. The underlined words are: 'nächste Frage', 'bitte', 'Haben Sie', 'hm', 'hm', 'Das ist mein Handy', 'Ja ihr habt ja jetzt erzählt, wie es am Anfang war', 'Wie ist es mittlerweile für euch, wie ist jetzt euer Eindruck von Deutschland', 'Perfekt', 'Das ist bestes, was mit uns passieren könnte', 'Wir sind echt sehr sehr zufrieden', 'hm', 'Wir mögen dieses Land', 'Wir wollen nicht zurück oder nach irgendwelches noch Land', 'Nein', 'Wir haben uns hier gefunden, also ich, für mich kann ich sagen ich bin daheim', 'Dahoam, oder wie sagt man das, ja', 'Dahoam', 'Dahoam, ja, ja, Genau', 'Ja, mir gefällt auch, ich habe über Deutschland geträumt. Ich wohne da', 'Ne', 'Und is, is es in etwa so wie Du es dir vorstellst hast, oder', 'Nein, ah, ich habe mir über deutsche Leute gedacht, zum Beispiel, gibt es einige intelligente, zum Beispiel freundlich wie in Kroatien, zum Beispiel gewohnt, aber anders, Das sind nicht, nicht Deutsche, dass sind Leute wir sagen, dass kann man nicht an der Nationalität ahn', 'Nu Leute, ja, Gibt es freundlich, gibt es unfreundlich und ja, aber ich habe gedacht, ein bisschen mehr freundlich müssen Deutschen sein', 'Der meint vielleicht, wir spüren, dass, an uns sozusagen, dass wir Ausländer sind und das kommt ein bisschen von anderen Leuten. Leute die nicht Ausländer mögen es gibt ja in jede Land, in jedem Land gibts solche Leute. Und da gibts ja auch die nicht zum Beispiel Russen mögen, es gibt auch. Und man spürt das. Es gibt sowas auch', 'Nun, nein, zum Beispiel ich verstehe einige Wörter nicht und das war schwer. Ich habe Führerschein gemacht, bestanden, aber Lehrerin war Oh du verstehst nicht, du verstehst. Ich habe gesagt ich lerne. Ich wohne nur in Deutschland ein Jahr. Was, muss ich perfekt sprechen, oder was. Ja, ich werde weiter lernen, ich freundlich und so weiter. Ich zum Beispiel, ich helfe immer. Wenn man braucht zum Beispiel braucht im Fitnessstudio Nachbar ich habe alles getragen. Wir sind nach Bar ähm zum Bank zum Beispiel gefahren und wir haben alles geholfen, aber andere zum Beispiel Hilfe Andere Oh Sie sprechen nicht perfekt. Tschüss', 'Ich glaube, man wird so behandelt, also wenn man hört wie wir auf Deutsch reden, ja. Und dann bildet man so ein Bild im Kopf, ja. Also, der redet nicht so gut Deutsch, vielleicht der ist dumm oder so oder keine Ahnung. Und so ein Eindruck war von uns vielleicht auf am Anfang. Und das ist ja vielleicht bis jetzt noch so, aber das heißt nicht, dass wir dumme Leute oder so, also wir haben ja auch die Schule besucht, wir können auch Schreiben, Lesen und wir können auch, wir sind auch in andere Richtungen gut, vielleicht wie andere Leute oder so, aber das ist das Problem. Man, man...', 'hm', 'Man macht den ersten Eindruck von dem wann man sieht und wenn man was hört von dem Mensch. Und das ist natürlich, glaube ich, war der Fall, ja. Weil wir haben natürlich im Jobcenter oder so, wir haben, wir könnten nicht so gut reden, aber irgendwie, ich hab glaube ich, ich kann sagen, dass ich Deutsch gelernt habe, so, ich hab nur einfach gesagt, was ich denke. Es war mir egal, wie ich den Satz baue. Bei mir war Nomen am Ende, Verb am Anfang, das war mir egal. <https://portal.oral-history.digital/mfi/de/interviews/mfi/>

Abb. 25: Textsuche mit dem höchsten Chunk zu Topic 44, Interviews des Museums Friedland. [Screenshot: Philipp Bayerschmidt 2026]

Bei den hier Interviewten handelt es sich um zwei Personen, die aus Kasachstan nach Deutschland migriert sind. Die Textpassage zeigt sehr schön, wie die interviewten Personen sich in Deutschland heimisch fühlen und sogar den dialektalen Ausdruck »Dahoam« verwenden, um ihre Verbindung zum Ankunftsland zu unterstreichen. [52]

Für den tatsächlichen Analyseprozess sind die Overview-Ansicht und die zahlreichen Verknüpfungen der Analyseebenen sehr hilfreich. Für die vorangegangenen Teilschritte wurden für eine deutlichere Darstellung die Einzelansichten der Elemente der Overview-Ansicht gewählt. [53]

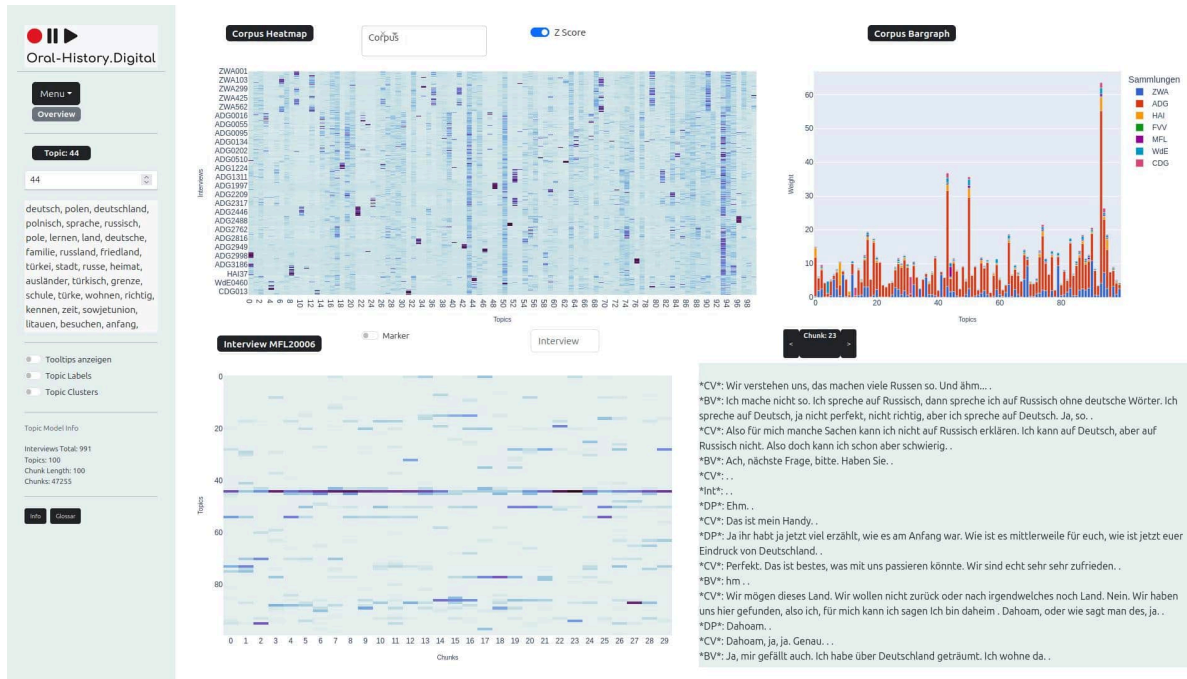


Abb. 26: Overview-Ansicht. [Screenshot: Philipp Bayerschmidt 2026]

Die »Chunk-Analyse« ist nun für das weitere Vorgehen von zentraler Bedeutung, da sie die Ergebnisse noch einmal nach einer anderen Logik sortiert, die für die Forschung relevant ist. Denn vordergründig werden Interviewpassagen, in denen über Heimat und Beheimatungsprozesse gesprochen wird, gesucht. Sollte dies beispielsweise nur in ein bis zwei Chunks stattfinden, würde das in der Korpus-Heatmap unter Umständen übersehen werden, da sie in der durchschnittlichen Topic-Verteilung im Interview untergehen. Deshalb ist die Kombination der verschiedenen Ergebnisebenen in der »Chunk-Analyse« sinnvoll, weil sie die Chunks als eigenständige Dokumente nutzt. Für die nächste Analyse wurde nach allen Chunks gesucht, deren Wert über 0.1 liegt, abhängig davon, dass der höchste Wert bei 0.5 lag und ein möglichst breites Spektrum abgebildet werden sollte. Der Grenzwert konnte deshalb festgelegt werden, weil bei vorherigen Analysen von einzelnen Textstellen auch passende Inhalte gefunden wurden, bei denen Werte von knapp über 0.1 lagen.

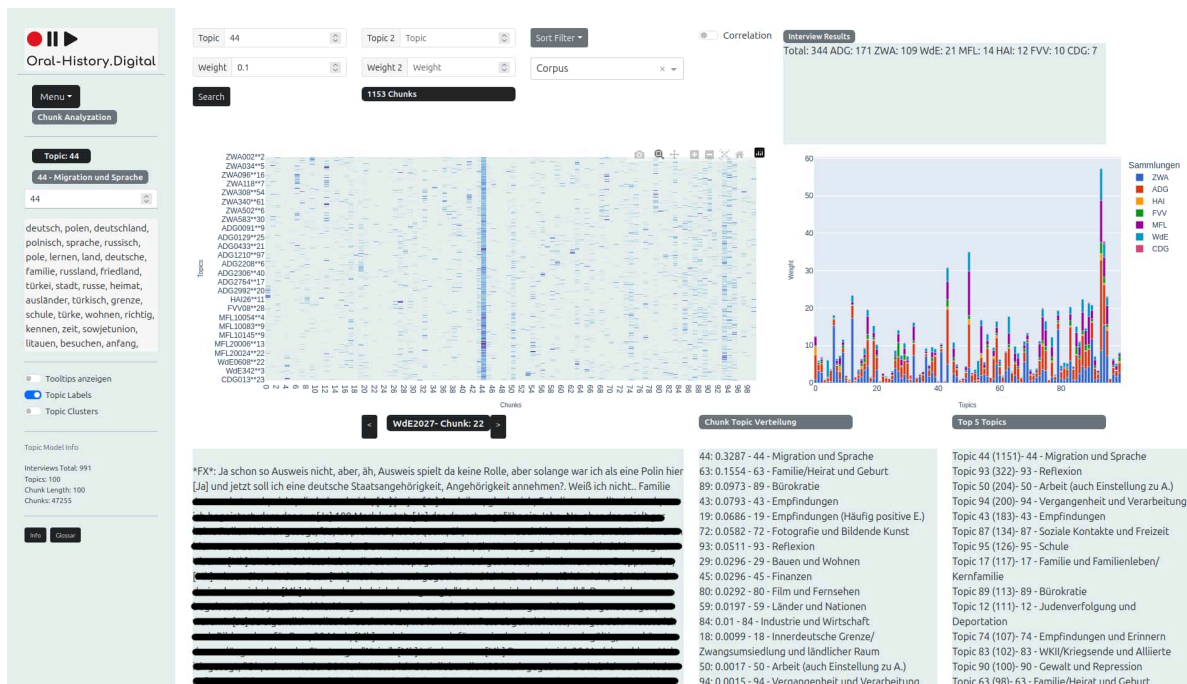


Abb. 27: Chunk-Analyzation, Interviewausschnitt FZH / WdE 2027. [Screenshot: Philipp Bayerschmidt 2026]

Das Ergebnis zeigt, dass es für Topic 44 insgesamt 1.153 Chunks mit einem Topic-Gewicht von über 0.1 gibt, die sich aus insgesamt 344 Interviews zusammensetzen. Die Anzeige oben rechts gibt noch eine genauere Aufteilung der Archive, aus denen die Interviews stammen, an. Diese mithilfe des Dashboards gefundenen Chunks können anschließend in einem close reading weiter ausgewertet werden und die Ergebnisse in die Topic-Korrelationen eingebunden werden.

[55]

## 6. Fazit

Die bisher bestehenden visuellen Analyse-Tools bieten in Ansätzen gute Möglichkeiten, Topic-Modeling-Ergebnisse zu analysieren, jedoch haben sie einige entscheidende Mängel. So können zum einen nicht immer alle Analyseebenen angesteuert werden, sondern meist nur eine eingeschränkte Auswahl, und zum anderen ist die gleichzeitige Darstellung der Ebenen sowie die interaktive Verknüpfung nicht immer ausreichend gegeben. Zudem sind einige der Angebote nicht open source zugänglich und besitzen eingeschränkte Import-Formate. Ein Grundproblem besteht in der Varianz und den Anforderungen verschiedener Quellengattungen und Forschungskorpora, die unterschiedliche Bedürfnisse an Topic-Modeling-Pipelines und Analyse-Tools stellen. Zusätzlich sind die notwendigen Analyseebenen und deren Kombination nicht nur von der Quellengattung, sondern auch von der jeweiligen Fragestellung abhängig.

[56]

Das OHTM-Dashboard bietet eine umfangreiche visuelle Analysemöglichkeit, die Zugriff auf nahezu alle vorhandenen Analyseebenen bietet und diese interaktiv miteinander kombiniert. Dies ermöglicht einen intuitiven Zugang, der sich Ebene für Ebene verfeinern lässt. Mithilfe des OHTM-Dashboards ist es möglich, das Korpus aus lebensgeschichtlichen Interviews auf ein gesuchtes Topic hin zu analysieren. Dabei ist das Dashboard nicht nur während der Analyse anwendbar, sondern auch bereits zuvor während der Topic-Modeling-Berechnung zur Validierung und zum Vergleich von verschiedenen Modellen, um Parameter zu bestimmen. Auch wenn das Dashboard durch die Ausgabe von Timecodes und einem Direktlink zum Interview im Archiv auf lebensgeschichtliche Interviews aus dem OH.D-Archiv spezialisiert ist, kann es ohne Probleme mit reinen TXT-Dateien genutzt werden. Denn das Dashboard funktioniert auf Grundlage des OHTM-Files, das durch die OHTM-Pipeline erstellt wird, die neben den Transkript-CSV-Dateien auch TXT-

[57]

Dateien verarbeiten kann. Durch den Open-Source-Zugang und einen modularen Aufbau ist sowohl die Pipeline als auch das Dashboard individuell anpassbar. Durch die Kombination des OHTM-Files und des Dashboards ist es darüber hinaus möglich, Forschungsergebnisse anderer Forscher\*innen zu evaluieren, indem Zugriff auf das OHTM-File gewährt wird und dieses mit dem Dashboard geöffnet werden kann.

## Bibliografie

- Jonathan Armoza: Topic Words in Context (TWiC), In: Jonathan Armoza: jarmoza. GitHub. 2017. Datenset. [\[online\]](#)
- Philipp Bayerschmidt: OHTM, an Oral History Topic Modeling Pipeline. In: Philipp Bayerschmidt: bayerschphi. GitHub. 2026. Datenset. [\[online\]](#)
- Philipp Bayerschmidt: OHTM-Dashboard. In: Oral History Digital. GitHub. 2025–2026. Datenset. [\[online\]](#)
- Philipp Bayerschmidt / Dennis Möbus: Leben gelabelt. Computergestützte Inhaltsverzeichnisse für Oral-History-Interviews. In: BIOS – Zeitschrift für Biographieforschung, Oral History und Lebensverlaufanalysen 38 (2025), H. 1+2, S. 83–104. [\[Nachweis im GVK\]](#)
- Beate Binder: Heimat als Begriff der Gegenwartsanalyse. Gefühle der Zugehörigkeit und soziale Imaginationen in der Auseinandersetzung um Einwanderung. In: Zeitschrift für Volkskunde 104 (2008), H. 1, S. 1–17. [\[Nachweis im GVK\]](#)
- David Meir Blei / Andrew Y Ng / Michael I. Jordan: Latent Dirichlet Allocation. In: Journal of Machine Learning Research 3 (2003), H. 4–5, S. 993–1022. [\[online\]](#) [\[Nachweis im GVK\]](#)
- David Meir Blei: Probabilistic Topic Models. In: Association for Computing Machinery (Hg.): Communications of the ACM 55 (2012), H. 4, S. 77–84. HTML. DOI: [10.1145/2133806.2133826](#)
- Allison Chaney / David Meir Blei: Visualizing Topic Models. In: ICWSM 2012. 6. International AAAI Conference on Weblogs and Social Media. Tagungsband (Dublin, 04.–07.06.2012). Dublin 2012, S. 419–422. HTML. DOI: [10.1609/icwsm.v6i1.14321](#)
- Jason Chuang / Christopher D. Manning / Jeffrey Heer: Termite: Visualization Techniques for Assessing Textual Topic Models. In: Genny Tortora / Stefano Levialdi / Maurizio Tucci (Hg.): AVI '12. International Working Conference on Advanced Visual Interfaces. Tagungsband (Capri, Italien, 21.–25.05.2012). New York 2012, S. 74–77. DOI: [10.1145/2254556.2254572](#)
- Marko Demantowsky: Was soll das bloß mit dieser Heimat. In: Jan Matthias Hoffrogge / Martin Schlutow / Max Twickler (Hg.): Geschichtsbewusstsein in der Gesellschaft. Festschrift für Bernd Schönemann zum 65. Geburtstag. Unter Mitarbeit von Bernd Schönemann (= Forum Historisches Lernen). Frankfurt / Main 2021, S. 170–188. [\[Nachweis im GVK\]](#)
- Lina Franken / Dennis Möbus: Mensch und Maschine als Team. Exploratives Topic Modeling und manuelle Annotation in der qualitativen Sozialforschung. In: Zeitschrift für digitale Geisteswissenschaften 9 (2024), 23.05.2024. HTML / XML / PDF. DOI: [10.17175/2024\\_003](#)
- Tobias Hodel / Dennis Möbus / Ina Serif: Von Inferenzen und Differenzen. Ein Vergleich von *Topic-Modeling-Engines* auf Grundlage historischer Korpora. In: Selin Gerlek / Sarah Kissler / Thorben Mämecke / Dennis Möbus (Hg.): Von Menschen und Maschinen – Mensch-Maschine-Interaktion in digitalen Kulturen. Hagen 2022, S. 185–209. PDF. DOI: [10.57813/20220623-153139-0](#)
- Jens Jäger: Heimat. In: zeitgeschichte DIGITAL. 09.11.2017. HTML. DOI: [10.14765/zzf.dok.2.1113.v1](#)
- Stefan Jänicke / Greta Franzini / Muhammad Faisal Cheema / Gerek Scheuermann: Visual Text Analysis in Digital Humanities. In: Computer Graphics Forum 36 (2017), H. 6, S. 226–250. DOI: [10.1111/cgf.12873](#)
- Christian Kahmann / Andreas Niekler / Gregor Wiedemann: Application of the Interactive Leipzig Corpus Miner as a Generic Research Platform for the Use in the Social Sciences. arXiv. 06.10.2021. DOI: [10.48550/arXiv.2110.02708](#)
- Benjamin Krautter: The Scales of (Computational) Literary Studies. Martin Mueller's Concept of Scalable Reading in Theory and Practice. In: Florentina Armaselu / Andreas Fickers (Hg.): Zoomland. Exploring Scale in Digital History and Humanities (= Studies in Digital History and Hermeneutics, 7). Basel u. a. 2024, S. 261–286. DOI: [10.1515/9783111317779-011](#)
- Svenja Kück: Heimat und Migration. Ein transdisziplinärer Ansatz anhand biographischer Interviews mit geflüchteten Menschen in Deutschland (= Sozial- und Kulturgeographie, 43). Bielefeld 2022. [\[Nachweis im GVK\]](#)
- Huan Liu / Yubo Tao / Yining Qiu / Wenda Huang / Hai Lin: Visual Exploration of Software Evolution via Topic Modeling. In: Journal of Visualization 24 (2021), H. 4, S. 827–844. PDF / HTML. DOI: [10.1007/s12650-020-00739-7](#)
- Beate Mitzscherlich: Heimat als subjektive Konstruktion. Beheimatung als aktiver Prozess. In: Edoardo Costadura / Klaus Ries / Christiane Wiesenfeldt (Hg.): Heimat global. Modelle, Praxen und Medien der Heimatkonstruktion (= Edition Kulturwissenschaft, 188). Bielefeld 2019, S. 183–196. [\[Nachweis im GVK\]](#)
- Dennis Möbus: Holleriths Vermächtnis – ein Beitrag zur Geschichte von Frauen in der EDV. Topic Modeling als Methode digitaler Sekundäranalyse. In: BIOS – Zeitschrift für Biographieforschung, Oral History und Lebensverlaufanalysen 33 (2020), H. 2, S. 162–180. [\[Nachweis im GVK\]](#)
- Dennis Möbus: Interview Chronology Analysis (ICA): Verläufe von (lebensgeschichtlichen) Interviews visuell analysieren. In: Nils Reiter / Thomas Haider / Daniel Kababgi / Hendrik Buschmeier (Hg.): DHd 2025. Under Construction. 11. Jahrestagung des Verbands Digital Humanities im deutschsprachigen Raum. Konferenzabstracts (Bielefeld, 03.–06.03.2025). Bielefeld 2025, S. 469–471. PDF. DOI: [10.5281/zenodo.15269143](#)
- Dennis Möbus / Philipp Bayerschmidt / Tobias Hodel / Ina Serif: Topic Modeling für die Geschichtswissenschaft (= Zeitschrift für digitale Geisteswissenschaften / Working Paper, 5). Wolfenbüttel 2026. 17.06.2026. HTML / XML / PDF. DOI: [10.17175/wp\\_2026b](#)
- Valerie Müller / Christian Sieg / Lars Linsen: Uncertainty-Aware Topic Modeling Visualization. arXiv. 18.10.2021. DOI: [10.48550/arXiv.2110.09247](#)
- Anja Oesterheld: Geschichte der Heimat. Zur Genese ihrer Semantik in Literatur, Religion, Recht und Wissenschaft (= Studien und Texte Zur Sozialgeschichte der Literatur, 157). Berlin u. a. 2021. [\[Nachweis im GVK\]](#)
- Samuel Rönnqvist / Xiaolu Wang / Peter Sarlin: Interactive Visual Exploration of Topic Models Using Graphs. arXiv. 19.09.2014. Version 2 vom 27.11.2014. DOI: [10.48550/arXiv.1409.5623](#)
- Susanne Scharnowski: Heimat. Geschichte eines Missverständnisses. Darmstadt 2019. [\[Nachweis im GVK\]](#)
- Carson Sievert / Kenneth Shirley: LDAvis: A Method for Visualizing and Interpreting Topics. In: Jason Chuang / Spence Green / Marti Hearst / Jeffrey Heer / Philipp Koehn (Hg.): Interactive Language Learning, Visualization, and Interfaces. Workshop. Abstracts (Baltimore, US-MD, 27.06.2014). Baltimore, US-MD 2014, S. 63–70. DOI: [10.3115/v1/W14-31](#)
- Severin Simmler / Thorsten Vitt / Steffen Pielström: Topic Modeling with Interactive Visualizations in a GUI Tool. In: ADHO 2019. Complexities. Annual International Digital Humanities Conference of the Alliance of Digital Humanities Organizations. Konferenzabstracts (Utrecht, NL, 09.07.–12.07.2019). Utrecht 2019. DOI: [10.34894/ENV3TX](#)
- Severin Simmler / Thorsten Vitt / Steffen Pielström: Topics Explorer. In: DARIAH-DE (Hg.): Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften. Letzter Zugriff: 20.03.2026. HTML. [\[online\]](#)
- Katharina Zeppezauer-Wachauer: Distant Reading, Close Reading, Scalable Reading. In: KONDE Weißbuch. Hg. von Helmut Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt »Kompetenznetzwerk Digitale Edition«. 2021. Handle: [11471/562.50.71](#)

## Quellen

- A., Frau, Interview adg0001, 10.09.1981. Archiv »Deutsches Gedächtnis«. Forschungsdatenzentrum des Instituts für Geschichte und Biographie. [\[online\]](#)
- A., Johanna, Interview adg0002, 19.02.1982, 31.03.1982, 15.04.1981. Archiv »Deutsches Gedächtnis«. Forschungsdatenzentrum des Instituts für Geschichte und Biographie. [\[online\]](#)
- M., Elfriede, Interview adg1215, 22.11.1990. Archiv »Deutsches Gedächtnis«. Forschungsdatenzentrum des Instituts für Geschichte und Biographie. [\[online\]](#)
- K., Reinhard, Interview adg2443, 12.12.2002. Archiv »Deutsches Gedächtnis«. Forschungsdatenzentrum des Instituts für Geschichte und Biographie. [\[online\]](#)
- H., Margret, Interview adg2467, 8.10.2003. Archiv »Deutsches Gedächtnis«. Forschungsdatenzentrum des Instituts für Geschichte und Biographie. [\[online\]](#)
- H., Hans, Interview adg0021, 19.11.1980. Archiv »Deutsches Gedächtnis«. Forschungsdatenzentrum des Instituts für Geschichte und Biographie. [\[online\]](#)
- FZH-WdE 2027, Interview mit Frau X (Alias) am 14.03.1993. Interviewer Patrick Wanger.

## Abbildungs- und Tabellenverzeichnis

- Tab. 1: Topic-Wort-Verteilung des Topics 44.
- Tab. 2: Beispielhafte Topic-Dokument-Verteilung.
- Abb. 1: LDAvis mit der globalen Topic-Ansicht auf der linken Seite und den Wort-Balkendiagrammen auf der rechten Seite. [Aus: [Sievert / Shirley 2014](#), S. 64]
- Abb. 2: »termvOverview« der Topic Cloud. [Aus: [Rönnqvist et al. 2014](#), S. 3]
- Abb. 3: [Müller et al. 2021](#), S. 4]
- Abb. 4: Dokumentenansicht für das close reading. [Aus: [Müller et al. 2021](#), S. 5]
- Abb. 5: Philipp Bayerschmidt 2026]
- Abb. 6: Detaillierte Topic-Dokument-Verteilung für ein Dokument im DARIAH Topics Explorer. [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 7: Philipp Bayerschmidt 2026]
- Abb. 8: Validierungsoberfläche des iLCM. [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 9: Analyseansicht des TwiC. [Aus: [Armoza 2017](#)]
- Abb. 10: Navigationsmenü im Menüband. [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 11: Philipp Bayerschmidt 2026]
- Abb. 12: Topic-Wörter-Ansicht. [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 13: Philipp Bayerschmidt 2026]
- Abb. 14: Z-Score-bereinigte Heatmap. [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 15: Interview-Heatmap, [Interviewausschnitt adg0001](#). [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 16: Analyseebenen in der Overview-Ansicht, [Interviewausschnitt adg0002](#). [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 17: Tooltip-Anzeige im Dashboard, [Interviewausschnitt adg1215](#). [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 18: Label-Anzeige für die einzelnen Topics, [Interviewausschnitt adg2443](#). [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 19: Cluster-Anzeige, [Interviewausschnitt adg2467](#). [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 20: Textsuche in den Chunks, Interviews des Museums Friedland. [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 21: Chunk-Analyzation mit Chunk-Suche und weiteren Analyseebenen, [Interviewausschnitt adg0021](#). [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 22: Topic-Wörter-Ansicht mit markiertem Topic 44. [Screenshot: Philipp Bayerschmidt 2026]
- Tab. 3: Topic-Wort-Verteilung des Topics 44.
- Abb. 23: Balkendiagramm mit markiertem Topic 44. [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 24: Heatmap mit markiertem Topic 44. [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 25: Textsuche mit dem höchsten Chunk zu Topic 44, Interviews des Museums Friedland. [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 26: Overview-Ansicht. [Screenshot: Philipp Bayerschmidt 2026]
- Abb. 27: Chunk-Analyzation, [Interviewausschnitt FZH / WdE 2027](#). [Screenshot: Philipp Bayerschmidt 2026]