



UNIVERSITÄT  
LEIPZIG

# AI based Methods for Using Text as Data in the Social Sciences

iLCM - an interactive text mining environment for social and economic scientists

Leipzig, 12.12.2023

Dr. Christian Kahmann



## Agenda

- Part 1: Introduction to iLCM
- Part 2: Overview of Available Methods inside iLCM
- Part 3: Real World Application
  - Transnorms
  - Oral History Digital

# ILCM

ADP-1013-013

Task ID: 216

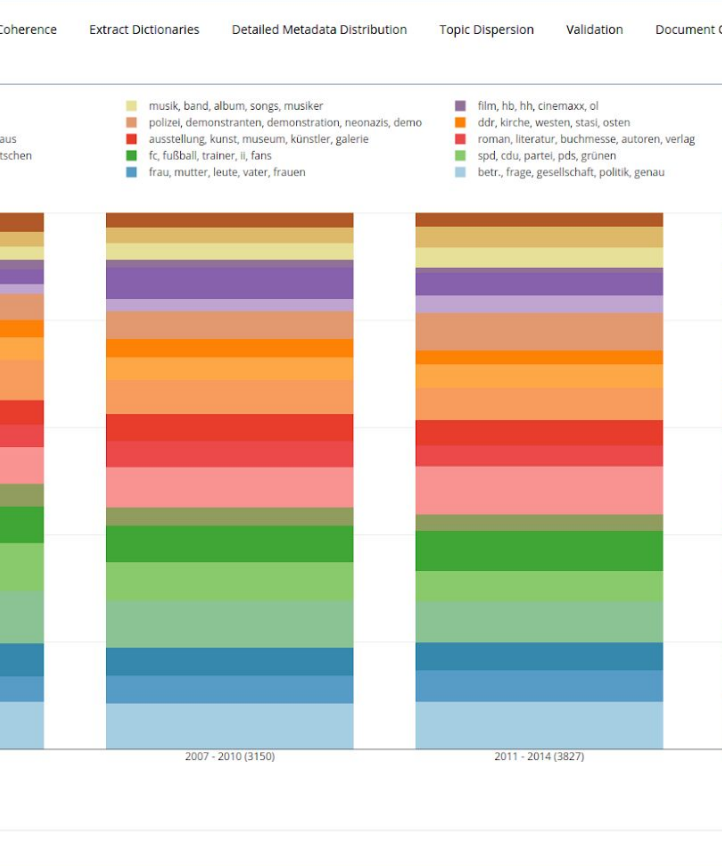
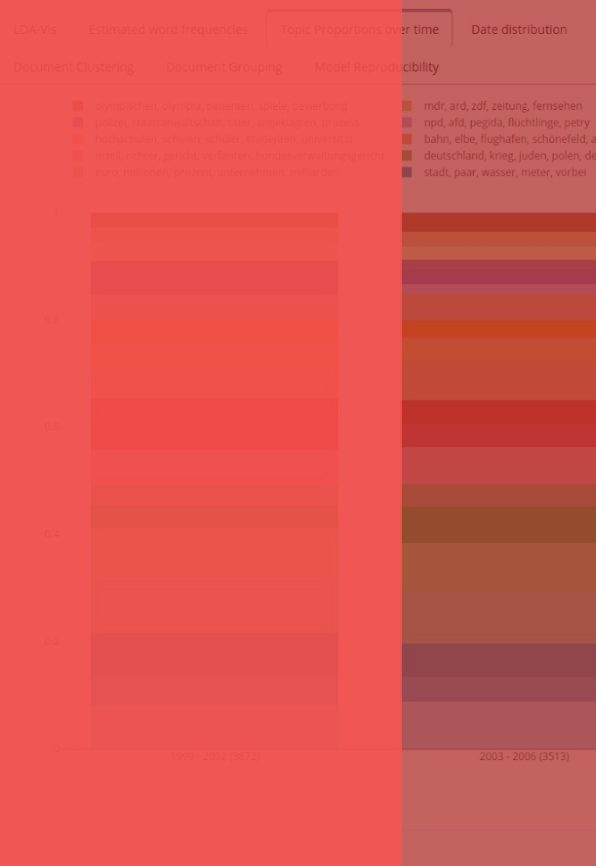
Lambda Value for Topic Labeling

Number of Words per Topic Label

By Date

Time Intervall

number of specified time intervals per used period





# ILCM

- Interactive Leipzig Corpus Miner
- DFG-funded project
- Joint project with GESIS
- Text mining infrastructure designed for use in the social sciences
- Support in the implementation of social science research questions

Goal: Enabling text mining for users with different prior knowledge and skills

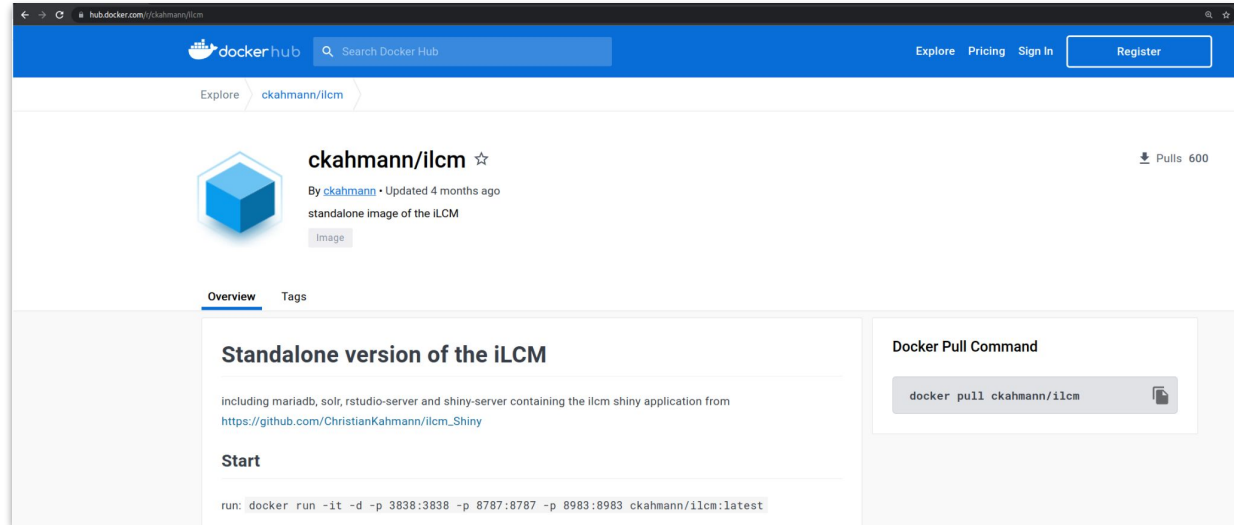
The screenshot displays the ILCM web interface. On the left is a navigation sidebar with options like 'Explorer', 'Collection Worker', 'Categories', 'Scripts', 'Import/Export', and 'Screenshot'. The main content area shows the title 'The Shawshank Redemption' and a list of metadata fields including 'id\_doc', 'id', 'dataset', 'score', 'date', 'token', 'language', 'en', 'Origin', and 'Director'. The central pane displays the text of the movie, with several phrases highlighted in red, indicating detected annotations. On the right side, there are two panels: 'Annotations' and 'Made Annotations'. The 'Annotations' panel shows a dropdown menu for 'Which Annotation Scheme?' with 'Violence vs Affection' selected, and a list of annotation types such as 'expression of affection', 'use of violence', and 'UNSET ANNOTATION'. The 'Made Annotations' panel shows a visual representation of the annotations with colored squares (red and blue) and a 'Remove' button.

## **ILCM**

- Specially developed for use in the social sciences
  - Allows the combination of quantitative and qualitative research approaches
  - Offers functions for both beginners and experienced users
  - Graphical user interface can be combined with script-based approaches
  - Generic orientation with focus on extensibility and customizability

## VIRTUALIZATION THROUGH DOCKER

- Image on Dockerhub
- → easy installation
- Ensuring the executability



The screenshot shows the Docker Hub interface for the image 'ckahmann/ilcm'. The page includes a search bar, navigation links (Explore, Pricing, Sign In, Register), and a star icon. The main content area features a blue cube icon, the image name 'ckahmann/ilcm', and a description: 'standalone image of the iLCM'. Below this, there are tabs for 'Overview' and 'Tags'. The 'Overview' tab is active, displaying the text 'Standalone version of the iLCM' and a link to the GitHub repository. A 'Start' section contains a Docker run command: `run: docker run -it -d -p 3838:3838 -p 8787:8787 -p 8983:8983 ckahmann/ilcm:latest`. On the right side, there is a 'Docker Pull Command' section with a button that says `docker pull ckahmann/ilcm`.

```
christian@workstation2: $ docker run -it -d ckahmann/ilcm:latest
```

<https://hub.docker.com/r/ckahmann/ilcm>

# GRAPHICAL USER INTERFACE

- All functions of the tool can be used via GUI

The screenshot displays the iLCM web interface. On the left is a navigation sidebar with options: Explorer, Collection Worker, Categories, Scripts, Import/Export, and a Screenshot button. The main content area is titled 'Document View' and shows the document 'The Shawshank Redemption'. A metadata panel on the left lists fields like id\_doc, id, dataset, score, title, date, token, language, Origin, and Director. The central text area contains the document's content with several phrases highlighted in red and blue, indicating annotations. On the right, there are two panels: 'Annotations' with a dropdown menu for 'Which Annotation Scheme?' (set to 'Violence vs Affection') and buttons for 'expression of affection', 'use of violence', and 'UNSET ANNOTATION'; and 'Made Annotations' showing a list of colored blocks and a 'Remove' button. The bottom left corner features the University of Leipzig logo.

## VARIABLE IMPORT INTERFACE

- Any text files or CSV file
- Mapping to internal format interactively within the GUI

Mapping

**Map title**

 automatic  
 title  
 section  
 date  
 uri  
 id  
 text

**Map date**

 automatic  
 title  
 section  
 date  
 uri  
 id  
 text

**Map body**

 title  
 section  
 date  
 uri  
 id  
 text

**Map uri**

 not required  
 title  
 section  
 date  
 uri  
 id  
 text

**Map section**

 not required  
 title  
 section  
 date  
 uri  
 id  
 text

**Map mde3**

 not required  
 title  
 section  
 date  
 uri  
 id  
 text

**Map mde4**

 not required  
 title  
 section  
 date  
 uri  
 id  
 text

**Map mde5**

 not required  
 title  
 section  
 date  
 uri  
 id  
 text

**Map mde6**

 not required  
 title  
 section  
 date  
 uri  
 id  
 text

**Language**  **dataset abbreviation**  **date format**

This dataset is not used yet. Feel free to specify your metadata

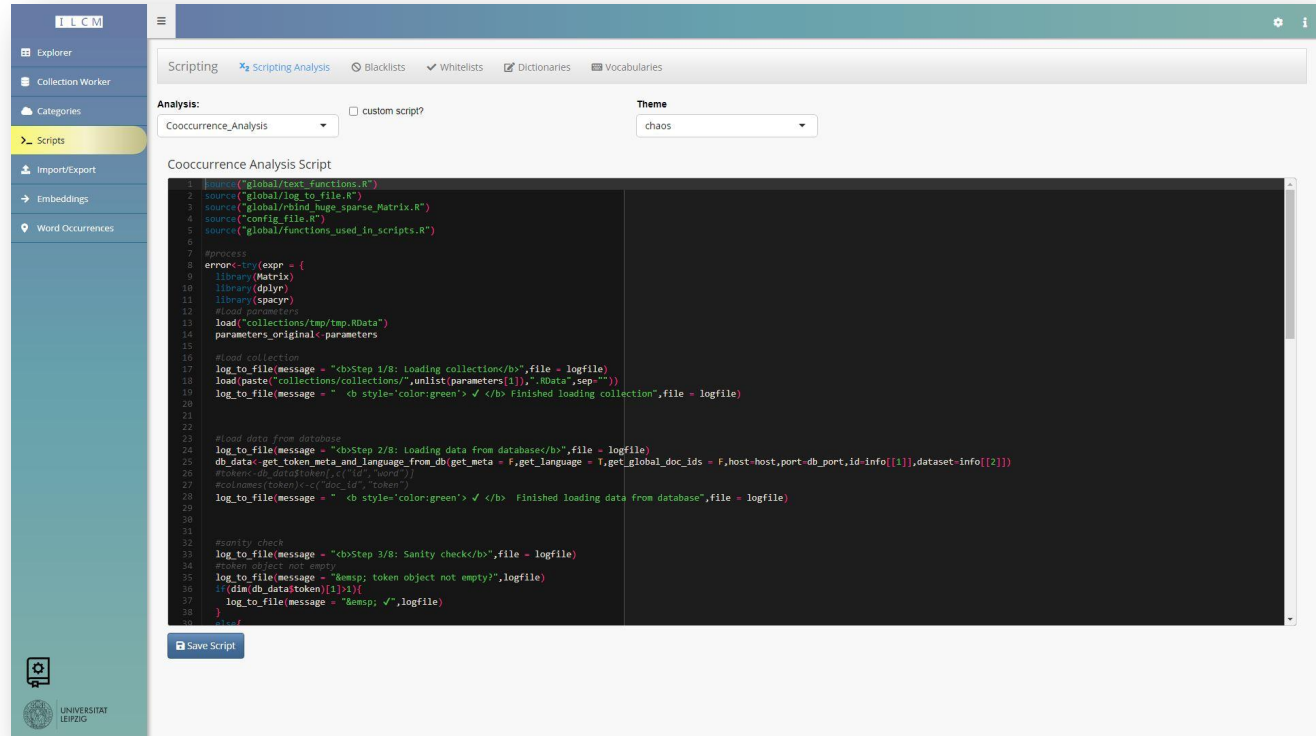
Resulting Metadata Import File

dataset	id_doc	title	body	date	token	language	uri	section
Gu	1	Did Bogle destroy Lauren Bacall?	Did Bogle destroy Lauren Bacall?. Even film stars rarely get to live out a screen romance. But Laur...	1999-01-01 17:09:00	2226	en	https://www.theguardian.com/film/1999/jan/01/features	Film
Gu	2	High hopes for cannabis	High hopes for cannabis. Britain's first commercial harvest of the prohibited drug cannabis sativa ...	1999-01-05 03:01:00	293	en	https://www.theguardian.com/business/1999/jan/05/7	Business
Gu	3	Western vanity wins in China by a reshaped	Western vanity wins in China by a reshaped nose. Lang Wenyu, a slight 26-year-old, lay on an operat...	1999-01-12 04:10:00	539	en	https://www.theguardian.com/world/1999/jan/12/2	World news



## SCRIPT-BASED ANALYSES

- Scripts can be customized within the GUI
- → *Custom Scripts*



## EXPORT OPTIONS

- All results & data exportable
- Standard formats
  - e.g. REFI standard
- → Further use in other software solutions



The screenshot displays the iLCM web interface. On the left is a sidebar with navigation options: Explorer, Collection Worker, Categories, Scripts, Import/Export (highlighted), Embeddings, and Word Occurrences. The main area is titled 'Importer Exporter' and has tabs for 'Collections', 'Results', 'Annotations', and 'All Files'. Under the 'Results' tab, there's an 'Analysis' section with checkboxes for classification, cooccurrence analysis, dictionary extraction, document deduplication, frequency extraction, sentiment analysis, topic model (checked), vector space representation, and volatility analysis. Below that is a 'Result sets' dropdown menu showing '76\_NDC PP nur EU\_2020-03-31 13:46:30'. A table titled 'Regular result files' lists various files with columns for filename, size in MB, creation time, and created by. At the bottom, there's a 'Special files to extract' section with buttons for 'theta' and 'phi'.

filename	size in MB	creation time	created by:	Download
data_TM.RData	3.57866	2020-03-31T11:47:06Z	rstudio	
documents_TM.RData	0.002075	2020-03-31T11:47:06Z	rstudio	
dtm_TM.RData	0.46388	2020-03-31T11:47:06Z	rstudio	
info.RData	0.005979	2020-03-31T11:47:06Z	rstudio	
meta_TM.RData	1.900124	2020-03-31T11:47:06Z	rstudio	
parameters.RData	0.000584	2020-03-31T11:47:06Z	rstudio	
vocab_2020-03-31.txt	0.086176	2020-03-31T11:47:06Z	rstudio	

# Methods



# INTERACTIVE DEDUPLICATION

Parameters

**Default strategy**

longest

**Threshold**

0,4

Reset user input

Save Collection

use iGraph-layout?

Help

duplicate free data

list of duplicates

Graph
Table

Document: 15283

Document Setting
Diff View

**Compare with:**

21210

Document: 15283	Document: 21210
<p>1 Studies using highly informative genetic markers (e.g. microsatellites) have revealed effective movement of genes through pollen grains from more distant sources into offspring than those inferred on the basis of physical pollen movement alone. In order to identify the factors that cause such effective pollen movement in tree populations analyses of the genetic structure of reproductive trees and parentage analyses for seedlings and saplings were conducted in a 110-ha population of <i>Aesculus turbinata</i> a tall deciduous tree species that is pollinated by insects and has large seeds that are dispersed by gravity and small mammals. 2. By analysing genotypes of seed coats which are identical to those of the seed parents and of leaves from seedlings pollen donors and seed parents were identified for seedlings. Differences in size and dispersal mode between pollen and seeds caused a significant difference in dispersal distances. Reflecting the restricted seed dispersal (25.3 m on average) there was significant positive relatedness between reproductive trees standing within 150 m of each other. 3. The percentage of self-pollinated offspring decreased from 8.3% at the seedling stage to 0.56% at the sapling stage. The disappearance of outcrossed</p>	<p>1 Studies using highly informative genetic markers (e.g. microsatellites) have revealed effective movement of genes through pollen grains from more distant sources into offspring than those inferred on the basis of physical pollen movement alone. In order to identify the factors that cause such effective pollen movement in tree populations analyses of the genetic structure of reproductive trees and parentage analyses for seedlings and saplings were conducted in a 0-ha population of <i>Aesculus turbinata</i> a tall deciduous tree species that is pollinated by insects and has large seeds that are dispersed by gravity and small mammals. 2. By analysing genotypes of seed coats which are identical to those of the seed parents and of leaves from seedlings pollen donors and seed parents were identified for seedlings. Differences in size and dispersal mode between pollen and seeds caused a significant difference in dispersal distances. Reflecting the restricted seed dispersal (25.3 m on average) there was significant positive relatedness between reproductive trees standing within 50 m of each other. 3. The percentage of self-pollinated offspring decreased from 8.3% at the seedling stage to 0.56% at the sapling stage. The disappearance of outcrossed</p>

Dismiss

## DOCUMENT RETRIEVAL

- Search via Solr Engine
- Inclusion of metadata possible
- → Creation of document collections

Simple Detailed Custom

Simple search. Terms may be combined with + # -  
+= AND  
#= OR  
- = NOT

**Keyword**

Simple Detailed Custom

**Detailed**

**Keyword**

Use raw text

**Date von:**

**Date bis:**

**Origin**

**Director**

**Cast**

**Genre**

**Wiki Page**

**Running Time**

**Number of Token:**

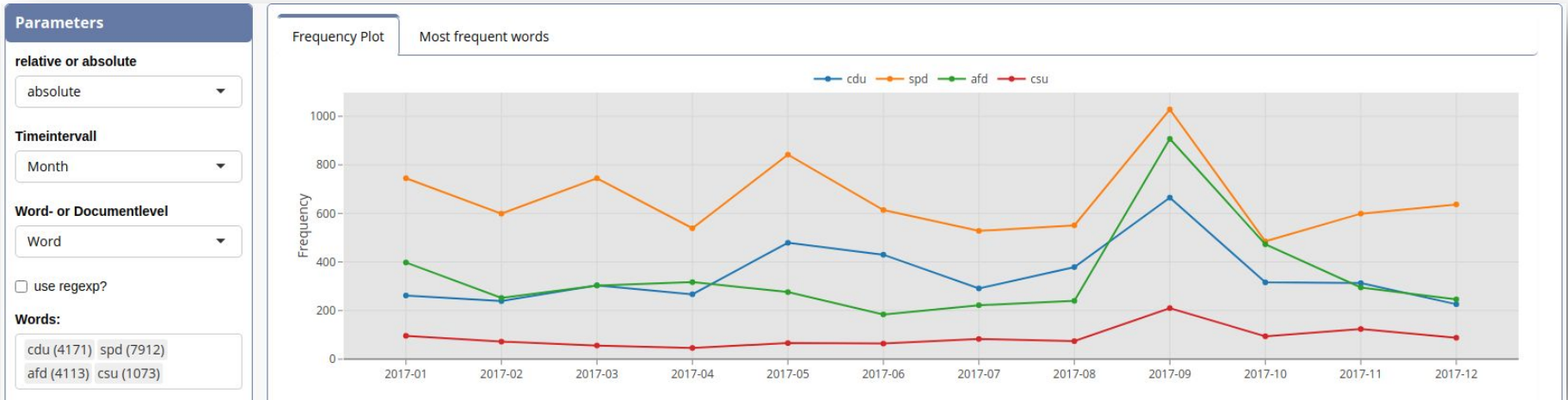
# DOCUMENT RETRIEVAL

The screenshot displays the iLCM search results page. The search term 'plane' is entered in the search bar. The results are shown in a table with columns: Year, dataset, title, date, language, token, id, score, Origin, Director, Cast, Genre, Wiki Page, Running Time, keyword\_and\_context, and Facets. A red circle highlights the table content. The facets on the right include: User raw text, Date von: 1971-03-12, Date bis: 2017-03-12, Origin: American, Director, Cast, Genre: action, Wiki Page, Running Time, Number of Tokens, and Save Collection.

Year	dataset	title	date	language	token	id	score	Origin	Director	Cast	Genre	Wiki Page	Running Time	keyword_and_context
2014	movies	Passenger 57	1992-03-12	en	735	6940521	12.505601	American	Kevin Hooks	Wesley Snipes, Bruce Payne, Tom Sizemore	action	https://en.wikipedia.org/wiki/Passenger_57	84	passengers, kill the FBI agents, release plane, and secure the plane by also shooting the captain, Celler, in
883	movies	Con Air	1997-03-12	en	774	6938890	12.255386	American	Simon West	Nicolas Cage, John Cusack, John Malkovich, Steve Buscemi, Ving Rhames, Mykelti Williamson, Rachael Ticotin	action	https://en.wikipedia.org/wiki/Con_Air	115	"Pinball" Parker incites a riot, releasing Orson and Damien Dog taking over the plane, planning to land
1300	movies	Executive Decision	1996-03-12	en	487	6939307	11.934199	American	Stuart Baird	Kurt Russell, Halle Berry, John Leguizamo, Oliver Platt, Joe Morton, David Suchet, Steven Seagal	action	https://en.wikipedia.org/wiki/Executive_Decision	133	Intelligence community, and DABPA engineer "Calli" join a team led by Travis to intercept the plane. After
1913	movies	I Spy	2002-03-12	en	676	6939920	11.752532	American	Betty Thomas	Owen Wilson, Eddie Murphy, Famke Janssen, Malcolm McDowell	action	https://en.wikipedia.org/wiki/I_Spy_(film)	97	"SwissMobile" plane sold to arms dealer Arnold Gundars. Gundars is sponsoring Midweight world boxing
1100	movies	Die Hard 2	1990-03-12	en	722	6939107	11.615754	American	Renny Harlin	Bruce Willis, Bonnie Bedelia, William Atherton, William Sadler	action	https://en.wikipedia.org/wiki/Die_Hard_2	124	plane so they can escape to another country with Esperance in tow, and warn the airport controllers not
2009	movies	Iron Eagle	1986-03-12	en	602	6940016	11.502123	American	John Flynn	Louis Gossett Jr., Jason Gedrick	action	https://en.wikipedia.org/wiki/Iron_Eagle	117	emerging battle, they take out three MiG-23 fighters and destroy an airliner, but Chapp's plane is damaged
4879	movies	Turbulence	1997-03-12	en	600	6942886	11.502123	American	Robert Butler	Ray Liotta, Lauren Holly, Hector Elizondo, Rachel Ticotin, Catherine Hicks, Brendan Gleeson	action	https://en.wikipedia.org/wiki/Turbulence_(1997_film)	100	capable of keeping the 747 from crashing. To make matters worse, the plane is heading for a Category 5



# FREQUENCY ANALYSIS



# SENTIMENT ANALYSIS

- Sentiment score as a scaling system for emotions in texts
- In iLCM based on predefined features
- German: SentiWS
- English: AFINN
- These features have weights in the interval [-1, 1]


```
"", "V1", "V2"
"1", "Abbau", "-0.058"
"2", "Abbruch", "-0.0048"
"3", "Abdankung", "-0.0048"
"4", "Abdämpfung", "-0.0048"
"5", "Abfall", "-0.0048"
"6", "Abfuhr", "-0.3367"
"7", "Abgrund", "-0.3465"
"8", "Abhängigkeit", "-0.3653"
"9", "Ablehnung", "-0.5118"
"10", "Ablenkung", "-0.0435"
"11", "Abnahme", "-0.0048"
"12", "Abneigung", "-0.0048"
"13", "Abnutzung", "-0.0048"
"14", "Abriss", "-0.0048"
"15", "Abrutsch", "-0.0048"
"16", "Abschaffung", "-0.058"
"17", "Abschreckung", "-0.0048"
"18", "Abschreibung", "-0.3345"
"19", "Abschuß", "-0.0048"
"20", "Abschwächung", "-0.1935"
```

**Parameters**

**Document:**

Seven (-0.31)

current plot data



Time Series   Document Length   Other meta data   **Validation**

Documents   [Top Documents](#)

Seven

Document Sentiment Score: **-0.309090909090909**

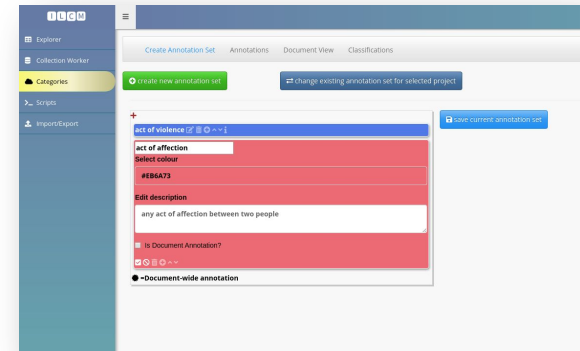
Soon - to - retire detective William Somerset is partnered with short - tempered but Idealistic David Mills , who has recently moved to the city with his wife Tracy . Tracy confides to Somerset that she is pregnant and has yet to tell Mills , as she is **unhappy** with the city and feels it is **no** place to raise a child . Somerset sympathizes , having had a similar situation with his ex - girlfriend many years earlier , and advises her to tell Mills only if she plans to keep the child . Somerset and Mills investigate a set of **murders inspired** by the seven deadly sins : a man **forced** to eat until his stomach ruptured , representing gluttony , and a defense attorney **killed** after a pound of flesh was taken from him , representing **greed** . Clues at the **murder** scenes lead them to a **suspect**'s apartment , where they find a third **victim** , a drug dealer and child molester , strapped to a bed , emaciated but **alive** , representing sloth . Daily photographs of the **victim** , taken over a year , show the crimes were planned far in advance . The detectives use library records to identify a John Doe and track him to his apartment . Doe **flees** and Mills gives chase . Doe turns to hold Mills at gunpoint for a moment before **escaping** . The apartment contains hundreds of notebooks revealing Doe 's psychopathy , as well as a clue to another **murder** . The detectives arrive too late to **stop** a man **forced** to **kill** a prostitute by raping her with a bladed strap - on , representing lust . They are told of a fifth **murder** , a model whose face has been mutilated by Doe ; she was given the option to call for **help** and be disfigured , or **commit suicide** by taking pills , representing pride . As Somerset and Mills return to the police station , Doe turns himself in , covered in the blood of an unidentified **victim** . Doe offers to take the detectives to the final two **victims** and confess to the **murders** , but only under specific terms , or he will plead **insanity** . Somerset is wary , but Mills **agrees** . The detectives follow Doe 's directions to a remote desert location . Within minutes , a delivery van approaches . Mills holds Doe at gunpoint while Somerset intercepts the driver , who has been instructed to bring a box to them . Doe taunts Mills by telling him how **jealous** he was of his life and Tracy . Somerset opens the box and **warns** Mills to stay back . Doe states that he **killed** Tracy , representing **envy** ; he also states that her head is in the box and that she was pregnant . Despite Somerset 's **warnings** , Mills shoots Doe , completing Doe 's last **murder** , representing wrath . Police converge and take the **devastated** Mills away .





# CLASSIFICATION

- Creation of codebooks
- Manual annotation of text references
- Active Learning
- Classification of unknown texts



Create Annotation Set Annotations Document View Classifications

Metadata

Part of Speech Tags

Named Entry Tags

id_doc	id	dataset	title	date	token	language
4587	6942594	movies	The Shawshank Redemption	1994-03-12	703	

## The Shawshank Redemption

In 1947 Portland, Maine, banker Andy Dufresne is convicted of **murdering his wife and her lover**, and sentenced to two consecutive life sentences at the Shawshank State Penitentiary. He is befriended by contraband smuggler, Ellis "Red" Redding, an inmate serving a life sentence. Red procures a rock hammer, and later a large poster of Rita Hayworth for Andy. Working in the prison laundry, Andy is regularly assaulted and raped by "the Sisters" and their leader, Bogs. In 1949, Andy overhears the captain of the guards, Byron Hadley, complaining about being taxed on an inheritance and offers to help him shelter the money legally. After an assault by the Sisters nearly kills Andy, Hadley beats Bogs severely. Left crippled, Bogs is transferred to another prison, and Andy is not attacked again. Warden Samuel Norton meets Andy and reassigns him to the prison library to assist elderly inmate Brooks Hatlen. Andy's new job is a pretext for him to begin managing financial matters for the prison employees. As time passes, the warden begins using him to handle matters for himself and a variety of people, including guards from other prisons. Andy begins writing weekly letters to the state legislature requesting funds to improve the prison's decaying library. Brooks is paroled in 1954 after serving 50 years, but he can not adjust to the outside world, and he commits suicide by hanging himself. Andy receives a library donation that includes a recording of The Marriage of Figaro. He plays an excerpt over the public address system, and is punished with solitary confinement. After his release from solitary, Andy explains that hope is what gets him through his time, a concept that Red dismisses. In 1963, Norton begins exploiting prison labor for public works, profiting by undercutting skilled labor costs and receiving bribes. Andy launders the money using the alias Randall Stephens. Tommy Williams is incarcerated for burglary in 1965. **Andy and Red befriend him**, and Andy helps him pass his GED exam. A year later, Tommy reveals to Red and Andy that an inmate at another prison claimed responsibility for the murders for which Andy was convicted. Andy approaches Norton with this information, but he refuses to listen and sends him back to solitary confinement when he mentions the money laundering. Norton has Hadley murder Tommy under the guise of an escape attempt. Andy declines to continue the laundering, but he relents after Norton threatens to burn the library, remove Andy's protection from the guards, and move him to worse conditions. After two months, Andy is released from solitary confinement, and he tells Red of his dream of living in Zihuatanejo, a Mexican coastal town. Red feels Andy is being unrealistic, but promises him that if he is ever released, he will visit a specific hayfield near Buxton, Maine, and retrieve a package Andy buried there. He worries about Andy's well-being, especially when he learns Andy asked another inmate to supply him with six feet (1.8 meters) of rope.

Annotations

Which Annotation Scheme?

Violence vs Affection

expression of affection i

use of violence i

UNSET ANNOTATION

Document-wide annotation

Save Annotations

Made Annotations

Remove

Evaluate Classification Evaluate Document

### expression of affection

Load new examples: Random Next

Save Dismiss

Charlie reconciles with his mother before telling his former love interest, Amelia, that he is still in love with her.

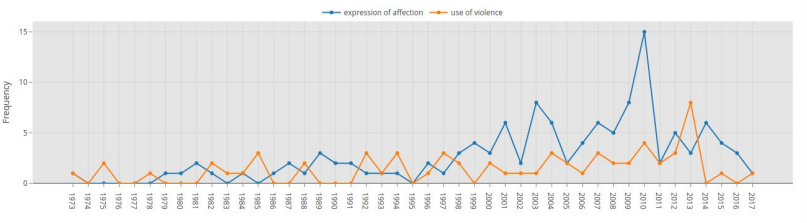
See document approve deny ignore

A comedy about "the powerlessness of being in love, how love devours the insides of a person like a deadly virus".

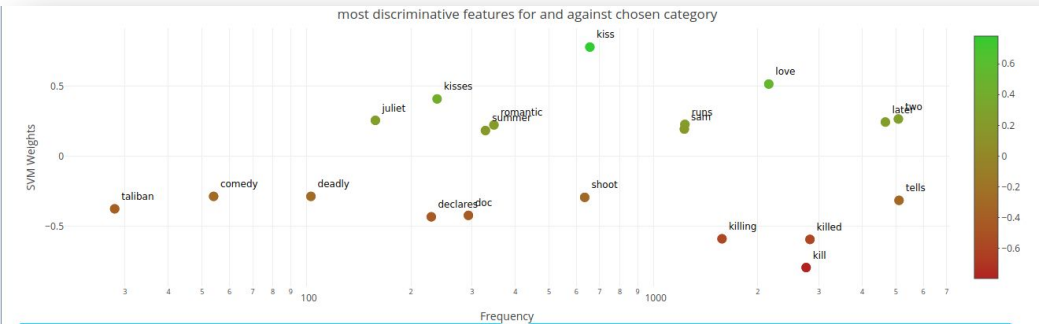
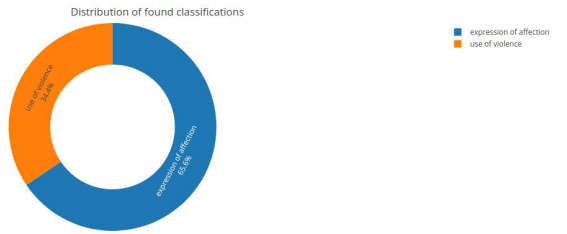
See document approve deny ignore

other category?  
 NEG  expression of affection

Date distribution Classifier performance Feature breakdown Validation



Download timeseries data



Pro features

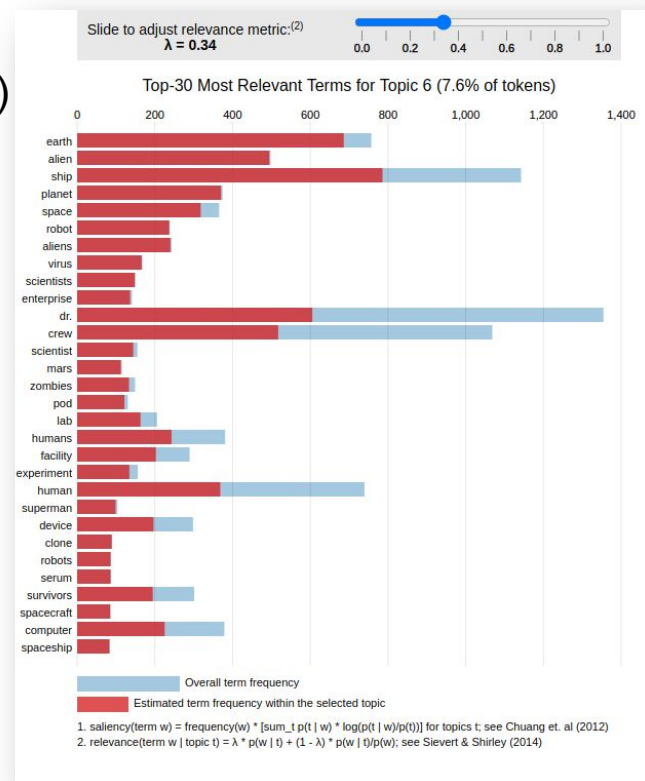
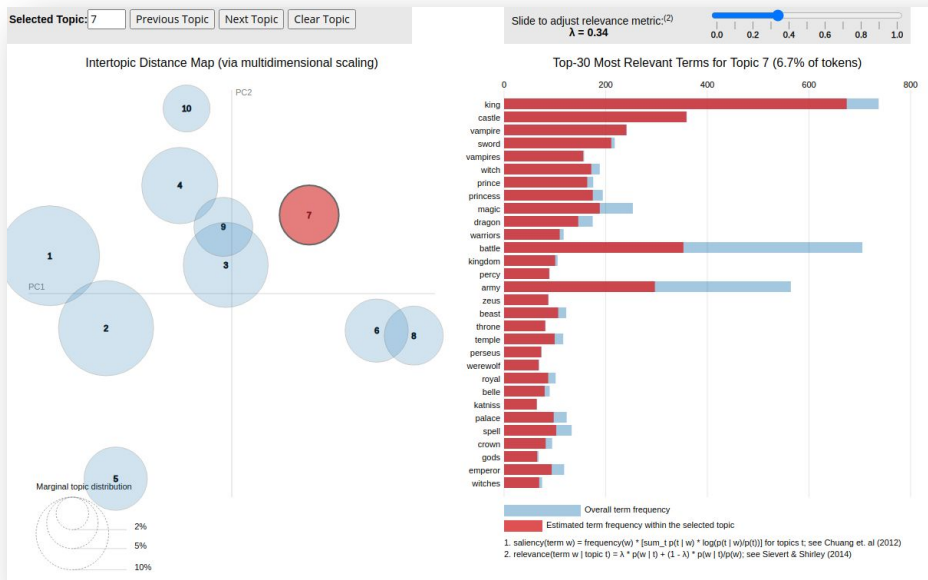
	weight	frequency
kiss	0.779186621636406	656
love	0.515130647789231	2151
kisses	0.409179429979813	238
two	0.265880462847006	5084
juliet	0.256141249219445	158
later	0.244054226552312	4663
runs	0.22747220277855	1233

Contra features

	weight	frequency
kill	-0.794295386250392	2756
killed	-0.593446744267932	2827
killling	-0.58852895504725	1577
declares	-0.432108105808401	229
doc	-0.421941086148722	293
taliban	-0.375090901987865	28
tells	-0.31453045439314	5106

# TOPIC MODELS

- Unsupervised classification into semantic areas (Topics)
- Documents: Distribution of Topics
- Topics: Distribution of Words



# TOPIC MODELS

iLCM
☰
⚙️

☰ Results
Details
Documents
Document View
Task Scheduler
My Tasks

Parameters

Task ID: 346

**Document selection:**

independently

by topic likelihood

**number of documents in selection**

**most relevant for which topic?**

**Document:**

(0.88) WALL-E (movies\_5006)

**Topic:**


⚙️

**Metadata**

title: WALL-E

date: 2008-03-12

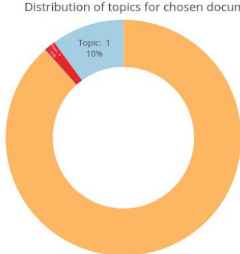
number of token: 524



LDA-Vis
Estimated word frequencies
Topic Proportions over time
Date distribution
Coherence
Extract Dictionaries
Detailed Metadata Distribution
Topic Dispersion
Validation
Document Comparison
Document Outlier


Document Clustering
Document Grouping
Model Reproducibility
Topic Topic Connection

Distribution of topics for chosen document



In 2805, **Earth** is an abandoned, dystopian **planet** covered in garbage, with its people evacuated by megacorporation Buy - N - Large on giant starliners . Bnl has left behind WALL - E **robot** trash compactors to clean up ; however , all have since stopped functioning , except one unit who has gained sentience and is able to stay active using spare parts from other units . One day , WALL - E discovers a healthy seedling , which he returns to his home . Later , an unmanned **spaceship** lands and **deploys** EVE **probe** to scan the **planet** . WALL - E is infatuated with EVE , who is initially hostile but gradually befriends him . When WALL - E brings EVE to his trailer and shows her the **plant** , however , she suddenly takes the **plant** and goes into standby mode . WALL - E , confused , unsuccessfully tries to reactivate her . The **ship** then returns to collect EVE , and with WALL - E clinging on , returns to its **mothership** , the starliner Axiom . The Axiom 's passengers have become obese and feeble due to microgravity and reliance on an **automated** lifestyle , including the **ship** 's current captain , McCrea , who leaves the **ship** under the control of the **robotic** **autopilot** , AUTO . EVE is taken to the bridge , with WALL - E tagging along . McCrea is unprepared for a positive **probe** response , but learns that placing EVE 's **plant** in the **ship** 's Holo - Detector for verification will trigger a hyperjump back to **Earth** so **humanity** can recolonize it . However , AUTO orders his **robotic** assistant GO-4 to steal the **plant** to prevent this from happening . EVE initially believes WALL - E is responsible for the **plant** 's disappearance . With the **plant** missing , EVE is deemed **faulty** and taken to Diagnostics . WALL - E mistakes the **procedure** as torture , and in intervening accidentally frees a group of **malfunctioning robots** and causes both EVE and himself to be **designated** as rogue **robots** . Frustrated , EVE takes WALL - E to an escape **pod** to send him home , but they are interrupted when GO-4 arrives with the **plant** , placing it in a **pod** set to self - **destruct** , which WALL - E enters just before it is jettisoned . WALL - E escapes , saving the **plant** , and he and EVE reconcile and celebrate with a dance in **space** around the Axiom . EVE brings the **plant** back to Captain McCrea , who watches EVE 's recordings of **Earth** and concludes that they have to return . However , AUTO refuses , revealing his own secret no - return directive A113 , issued to BnL autopilots after the corporation concluded in 2110 that the **planet** could not be saved . He mutinies , tasing WALL - E and **deactivating** EVE and throwing them both down the garbage chute , then detaining the Captain . EVE **automatically reactivates** and helps WALL - E bring the **plant** to the **ship** 's Holo - Detector chamber ; AUTO tries to close the chamber , crushing WALL - E when he struggles to keep it open , but Captain McCrea is able to **deactivate** him and destroy GO-4 , while EVE inserts the **plant** to **activate** the hyperjump . Having arrived back on **Earth** , EVE **repairs** and **reactivates** WALL - E , but finds that his memory has been **reset** and his personality is gone . Heartbroken , EVE gives WALL - E a farewell kiss , which sparks his memory back to life and restores his original personality . WALL - E and EVE reunite as the **humans** and **robots** of the Axiom begin to restore **Earth** and its **environment** .

Most relevant words for chosen topic



# Application Examples



Oral-History.Digital

## AREAS OF APPLICATION OF THE ILCM

- Faculty of Political Science, FU Berlin
  - Investigation of norms on a global and local level
- Center for Digital Systems (CEDIS), FU Berlin
  - Content exploration of the interview archives from the project Oral-History.Digital
- Helmholtz Center for Environmental Research
  - Project to investigate the history of pollination research
- Institute of Sociology, Leibniz University Hannover
  - Gender-related differences in funding applications and research topics
- German Evaluation Institute for Development Cooperation
  - Automatic geocoding in evaluation reports
- European Center for Press and Media Freedom
  - Prototype for press freedom monitor
- Saxon Academy of Sciences and Humanities
  - Shifts in the meaning of the term "sustainability"



Oral-History.Digital



## AREAS OF APPLICATION OF THE ILCM

- **Faculty of Political Science, FU Berlin**
  - Investigation of norms on a global and local level
- **Center for Digital Systems (CEDIS), FU Berlin**
  - Content exploration of the interview archives from the project **Oral-History.Digital**
- **Helmholtz Center for Environmental Research**
  - Project to investigate the history of pollination research
- **Institute of Sociology, Leibniz University Hannover**
  - Gender-related differences in funding applications and research topics
- **German Evaluation Institute for Development Cooperation**
  - Automatic geocoding in evaluation reports
- **European Center for Press and Media Freedom**
  - Prototype for press freedom monitor
- **Saxon Academy of Sciences and Humanities**
  - Shifts in the meaning of the term "sustainability"



Oral-History.Digital







## Projekt Transnorms

- Projekt at FU Berlin
- Topic: **Translation of global norms to the local level**
- Areas of application:
  - Climate change/climate protection
  - Human Rights
- Question: (How) can these different interpretations/translations be analyzed with the help of automatic procedures?
  
- <https://transnorms.eu>

## Data

- INDCs from 162 countries (as of December 2019)
  - Intended Nationally Determined Contributions
- NDCs from 183 countries (as of December 2019)
  - Nationally determined contribution: “NDCs embody efforts by each country to reduce national emissions and adapt to the impacts of climate change. The Paris Agreement (Article 4, paragraph 2) requires each Party to prepare, communicate and maintain successive nationally determined contributions (NDCs) that it intends to achieve.”



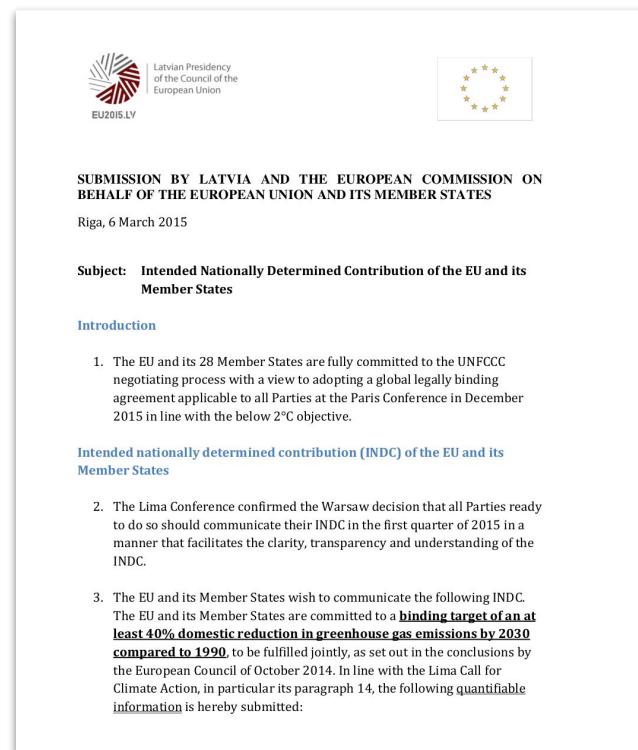
## Data

- <https://unfccc.int/NDCREG>
- Different in language, structure and size
- Examples:
  - Timor Leste: 32 pages, partly skewed scans with no OCR
  - EU: 5 pages, tables
  - Marshall Islands: 83 pages
  - Panama: 32 pages, spanish language



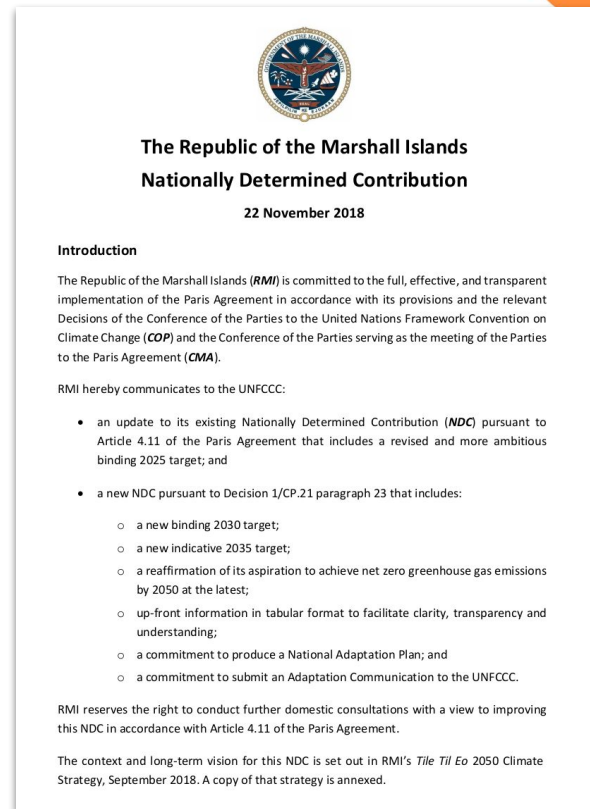
## Data

- <https://unfccc.int/NDCREG>
- Different in language, structure and size
- Examples:
  - Timor Leste: 32 pages, partly skewed scans with no OCR
  - EU: 5 pages, tables
  - Marshall Islands: 83 pages
  - Panama: 32 pages, spanish language



## Data

- <https://unfccc.int/NDCREG>
- Different in language, structure and size
- Examples:
  - Timor Leste: 32 pages, partly skewed scans with no OCR
  - EU: 5 pages, tables
  - Marshall Islands: 83 pages
  - Panama: 32 pages, spanish language



## Data

- <https://unfccc.int/NDCREG>
- Different in language, structure and size
- Examples:
  - Timor Leste: 32 pages, partly skewed scans with no OCR
  - EU: 5 pages, tables
  - Marshall Islands: 83 pages
  - Panama: 32 pages, spanish language



GOBIERNO DE LA REPÚBLICA DE  
**PANAMÁ**

**Contribución Nacionalmente  
Determinada a la Mitigación del  
Cambio Climático (NDC) de la  
Republica Panamá ante la  
Convención Marco de Naciones  
Unidas sobre Cambio Climático  
(CMNUCC)**



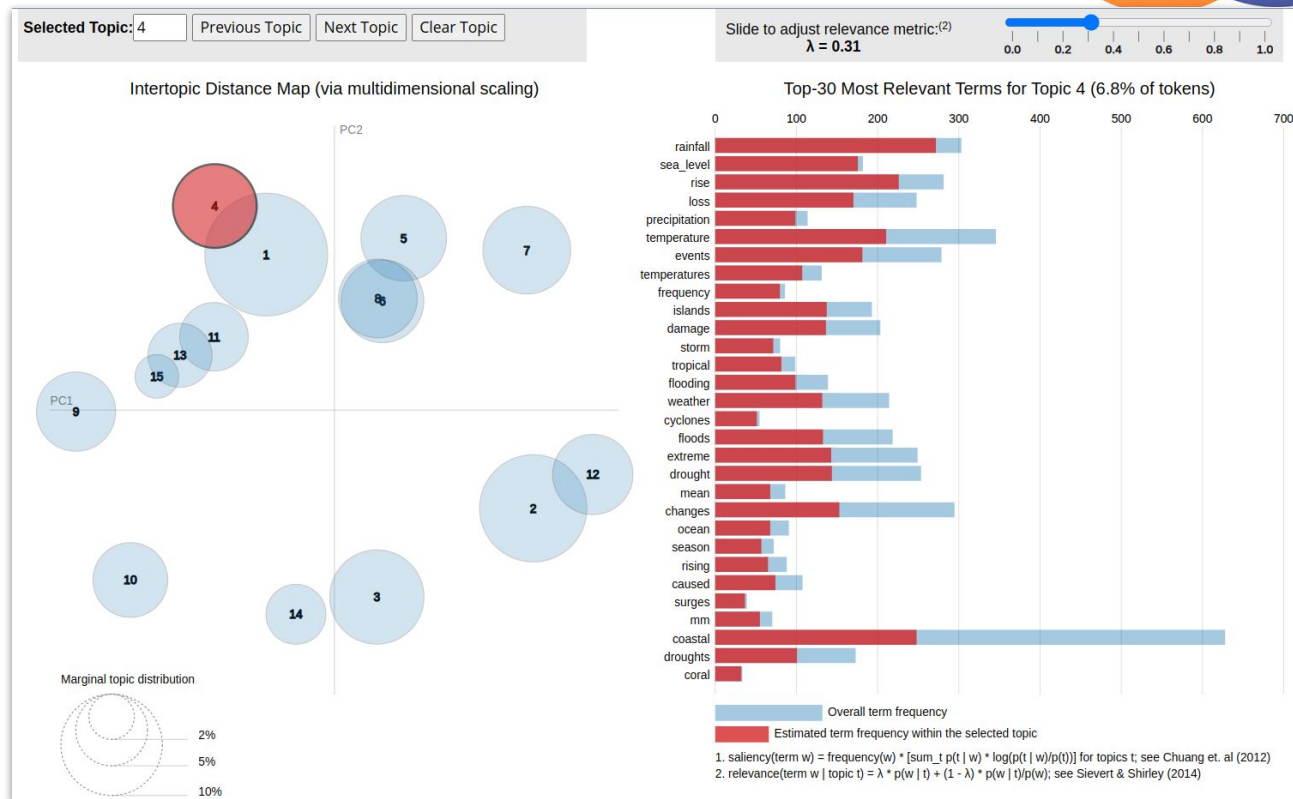
## Results

<b>Relative emission pledges</b> (134)	scenario, ghg, emissions, bau, conditional, baseline, mt, gg, unconditional, eq, <i>projection, reduction, lulucf, trend, scenarios, mitigation, year, intensity, equivalent, land_use, compared, calculated, tons, excluding, contribution</i>
<b>UNFCCC collaboration</b> (112)	convention, parties, agreement, unfccc, conference, nations, united, global, intended_nationally_determined_contribution, paris, article, cop, <i>differentiated, ambitious, decisions, common, submission, lima, effort, principles</i>
<b>Adaptation governance</b> (111)	climate_change, adaptation, resilience, risk, disaster, management, plan, climate, strengthen, resilient, <i>vulnerability, strategies, infrastructure, mainstreaming, policy, planning, security, impacts, integrated, response</i>
<b>Emission pledges &amp; markets</b> (82)	ipcc, accounting, target, per_cent, hfcs, pfcs, greenhouse_gas, removals, methodologies, hydrofluorocarbons, <i>guidelines, nitrous_oxide, commitment, economywide, mechanisms, parameter, timeframe, coverage, market, metric, credits, base</i>
<b>Renewable energy</b> (75)	energy, renewable, electricity, solar, generation, vehicles, fossil, grid, fuel, transport, pv, <i>demand, geothermal, energy_efficiency, fuels, diesel, landfill, wind, appliances, efficiency, heating, sources, road, mw, buildings</i>
<b>Economic growth</b> (72)	growth, economic, population, country, world, percent, poverty, today, challenges, gdp, <i>unemployment, economy, ranked, considerable, prospects, high, decade, gold, middle, prices, oil, problems, exports</i>
<b>Domestic governance</b> (71)	committee, national, ministries, preparation, ministry, process, coordination, council, cabinet, representatives, <i>stakeholders, technical, implementation, monitoring, sectoral, reporting, verification, institutional, evaluation, readiness, strategy</i>
<b>Forestry actions</b> (69)	forest, ha, wood, reforestation, afforestation, forests, stoves, deforestation, timber, hectares, <i>cookstoves, regeneration, planting, cooking, households, electrification, rice, agroforestry, firewood, sequestration, palm, plantations</i>
<b>Water-related disasters</b> (64)	rainfall, sea_level, rise, precipitation, loss, frequency, temperatures, events, temperature, islands, <i>storm, damage, tropical, flooding, cyclones, weather, floods, extreme, mean, drought, ocean, changes, season, rising, surges, caused, coastal</i>
<b>International support</b> (54)	financial, fund, funds, financing, required, support, needs, finance, capacity_building, funding, <i>needed, gcf, investment, donors, bilateral, gaps, skills, international, partners, technology_transfer, grants, cost, transfer</i>
<b>Water scarcity management</b> (51)	water, irrigation, diseases, varieties, crops, flood, livestock, crop, dams, surveillance, <i>catchment, food, pest, tolerant, disease, river, habitat, resistant, drinking, sanitation, soil, flow, early_warning, reuse, groundwater, drainage, fish</i>
<b>Land use &amp; anti-capitalism</b> (26)	beef, law, america, decree, native, lines, weight, grasslands, pncc, dated, <i>peoples, sovereignty, territorial, autonomous, justice, cuts, mother, territory, peatlands, compliance, shelter, articulation, rights, incorporation, capitalist</i>
<b>Conventional energy</b> (25)	mission, coal, launched, railway, gw, nuclear, rail, vehicle, pan, metro, <i>freight, standards, pricing, tonne, industries, clean, renewal, introduce, power, purchase, saving, stations, heater, railways, lowcarbon, encouraging, renovation, traffi</i>
<b>Pastoralism &amp; land management</b> (21)	drc, fca, slm, billion, inhabitant, conflicts, job, intensification, reinforcement, husbandry, <i>securing, cfa, hydroagricultural, oneoff, francs, populations, projects, creation, agroecological, orientations, nipa, animal</i>
<b>Post-war rural measures</b> (5)	charcoal, dam, lowlands, goe, acacia, war, moist, protein, moa, iucn, <i>species, project, federal, embankments, nomadic, napier, leaves, roots, civil, outputs, downstream, frankincense, cluster, highlands</i>



## Results

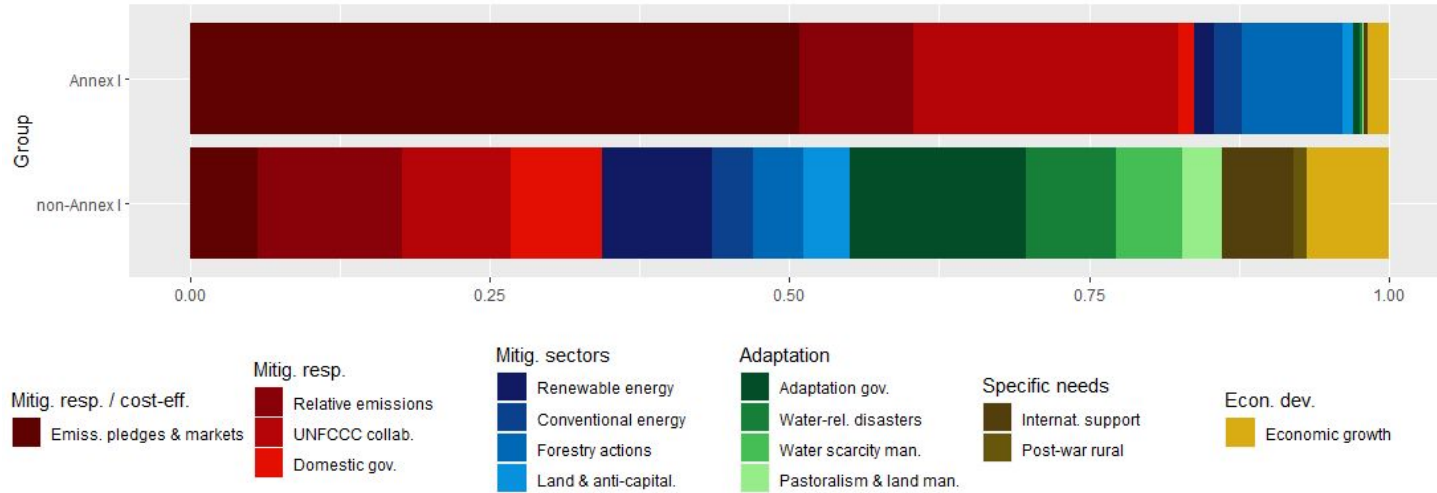
- “Water related disasters”



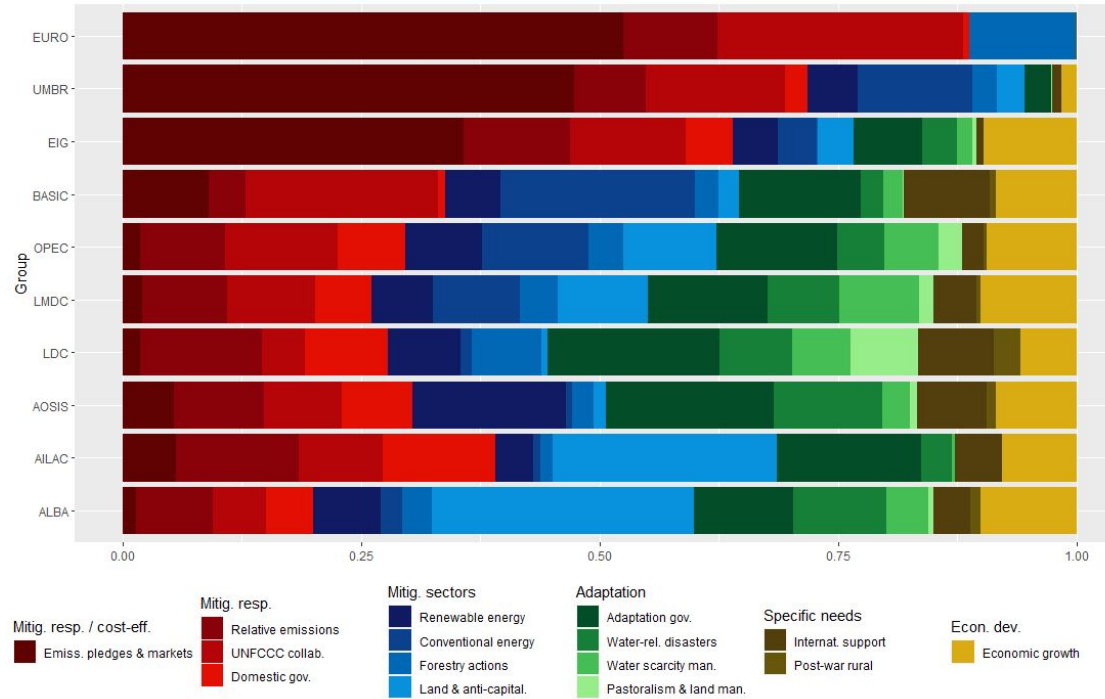




# Results



# Results





## Next Steps

- Analysis of Second NDCs → Change over time?
- Analysis of national Climate Documents → Change at local Level?
- Analysis of Carbon Closure Project → Change at Company Level?
- New focus: Human rights



## Projekt Oral History Digital

- Projekt at FU Berlin
- Topic: **Application of text mining to historical interview transcripts**
- Research question: How can the content analysis of historical interview transcripts be supported by the application of text mining?
  - Interview transcripts archive “Deutsches Gedächtnis” online
- <https://www.cedis.fu-berlin.de/services/projektentwicklung/aktuell/ohd/index.html>
- <https://deutsches-gedaechtnis.fernuni-hagen.de/de/interviews/adg0001>



## Projekt Oral History Digital

- Why do we need Text Mining?
  - Transcripts often very long (up to 10 hours)
  - Only individual issues relevant to research question
  - → Use Text Mining to identify relevant parts
  - Identification of trends beyond individual interviews
  
- Challenges:
  - New data format
  - So far a document is one continuous text
  - OHD: Line-by-line speech acts plus speaker information (interview format)
- → Adjustments



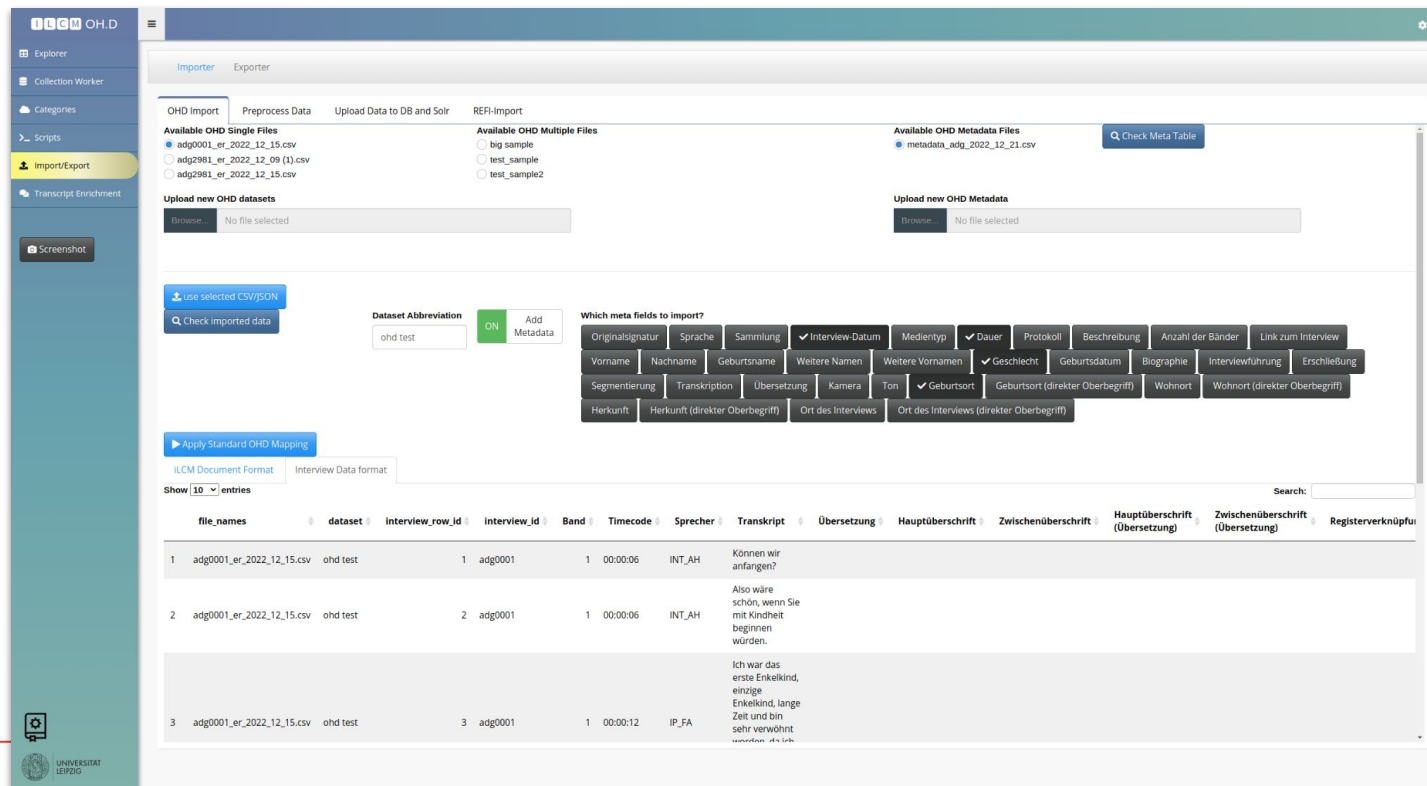


## Adjustments and method

- Adaptation to interview format
  - Filtering by speaker
  - Chunking based on speech acts
  - Delivery of results at the level of individual speaker acts
  - → Special input and export for OHD data type
- Application of topic modeling to find content focal points
- Subsequent marking of particularly relevant text sections with associated topics

# Results

## - OHD Import



The screenshot shows the 'OHD Import' interface in the iLCM environment. The interface is divided into several sections:

- Navigation Panel (Left):** Includes 'Explorer', 'Collection Worker', 'Categories', 'Scripts', 'Import/Export' (highlighted), 'Transcript Enrichment', and 'Screenshot'.
- Import/Export Section:**
  - Available OHD Single Files:** Three files are listed: 'adg0001\_er\_2022\_12\_15.csv', 'adg2981\_er\_2022\_12\_09 (1).csv', and 'adg2981\_er\_2022\_12\_15.csv'.
  - Available OHD Multiple Files:** Three radio buttons are present: 'big sample', 'test\_sample', and 'test\_sample2'.
  - Available OHD Metadata Files:** One file is listed: 'metadata\_adg\_2022\_12\_21.csv'.
  - Upload new OHD datasets:** A 'Browse...' button and 'No file selected' text.
  - Upload new OHD Metadata:** A 'Browse...' button and 'No file selected' text.
- Actions:**
  - 'Use selected CSV/JSON' (blue button)
  - 'Check imported data' (blue button)
  - 'Dataset Abbreviation' (input field with 'ohd test', 'ON' status, and 'Add Metadata' button)
  - 'Which meta fields to import?' (grid of checkboxes for various metadata fields)
  - 'Apply Standard OHD Mapping' (blue button)
  - 'ILCM Document Format' and 'Interview Data format' (tabs)
  - 'Show 10 entries' (dropdown)
- Table:**

file_names	dataset	interview_row_id	interview_id	Band	Timecode	Sprecher	Transkript	Übersetzung	Hauptüberschrift	Zwischenüberschrift	Hauptüberschrift (Übersetzung)	Zwischenüberschrift (Übersetzung)	Registerverknüpfung
adg0001_er_2022_12_15.csv	ohd test	1	adg0001	1	00:00:06	INT_AH	Können wir anfangen?						
adg0001_er_2022_12_15.csv	ohd test	2	adg0001	1	00:00:06	INT_AH	Also wäre schön, wenn Sie mit Kindheit beginnen würden.						
adg0001_er_2022_12_15.csv	ohd test	3	adg0001	1	00:00:12	IP_FA	Ich war das erste Enkelkind, einzige Enkelkind, lange Zeit und bin sehr verwöhnt werden, da ich						
- Footer:** 'UNIVERSITÄT LEIPZIG' logo and name.



# Results: Adaptation of document view to interview format

ILCM OH.D

Settings

**Explorer**

which corpus?

Collection Worker

Categories

Scripts

Import/Export

Transcript Enrichment

Screenshot

Search Results
Time Series
Document View
Facets
GeoExplorer

**Metadata**

Part of Speech Tags

Named Entity Tags

---

**id\_doc**

10

**dataset**

lusir sample

**title**

Interview: adg0101

**date**

2023-01-09

**language**

de

**token**

21708

**id**

61

**score**

1

**Interview-Datum**

23.9.1987

**Dauer**

## Interview: adg0101

Wie ich Ihnen schon sagte, ist also die erste Frage ... AD :

Also, Sie können mir aber zwischendurch auch mal Fragen stellen. Ja. Ja, aber die erste ist eben ganz pauschal Ihr ... Also können Sie aus ... Erst mal so ein Überblick ... Ihrem Leben erzählen. Ja, also nicht unbedingt ein Überblick. Sondern auch, also ... Mhm ... was Sie näher wollen. Also, das, das hört man jetzt, wenn ich kaue, ne? Muss mal wegtun. Nein, das macht aber nichts. Also, das ist ja nichts, also ... Also, geboren bin ich 1909, Januar 1909. Ja. In Prüm in der Eifel. Das ist ja auch ein historisch interessantes Städtchen. Ja. Da ist eine ..., war ein sehr schönes Gymnasium, eine alte Abtei. Ja. Und an diesem Gymnasium war mein Vater. Dadurch kamen wir nach Prüm. Ja. Da war mein Vater. Das nannte man da früher noch nicht Studienrat, sondern Oberlehrer. Aha. Ja, das war Oberlehrer. Und diese Zeit in Prüm, obwohl ich da nur also in den ..., aus den ersten fünf Jahren, bevor der Krieg begann, 1914 ... Ja. Ja. ... hab ich ja auch noch so allerhand Erinnerungen. Also, in so einer kleinen Stadt, also der Kaisergedanke war da, nicht. Ja. Ja. Es war das Kaiserreich. Und da mein Vater Beamter war, lebte man so ganz, eigentlich in diesem Nimbus Kaiserreich, nicht? Er war ein Angestellter des Kaisers, sozusagen. Ja. Ja. Und im Januar hatte mein Va ..., hatte der Kaiser Geburtstag, war Kaisers Geburtstag. Und da kam, mussten die Studienräte, also die damaligen Oberlehrer turnusmäßig Geburtstagsreden halten. Ja. Ja. Und dann mussten wir ganz still ... Wir Kinder mussten ganz still umhergehen, weil der Vater sich konzentrieren musste auf diese Rede. Und die ganze, der ganze Habitus war ausgerichtet auf den, also auf den preußischen Beamten. Ja. Ne. Alle Kinder trugen Matrosenkleider, das war ganz selbstverständlich. Ja. Ja. Ich erinnere mich, wie wir so auf der Bank, auf dem Sofa saßen, und betrachteten mit unserer Mutter die verschiedenen, äh, die, all die Kinder des Kaisers. Er hatte ja so furchtbar viele Kinder. Ja. Ja. Das jüngste war ein Mädchen. Und ich weiß noch, dass der Name Prinz Eitel Friedrich mir ungeheuren Eindruck gemacht hat. Einer war der Prinz Eitel Friedrich. Mhm. Diese Gesellschaft, das gesellschaftliche Gefüge war auch sehr stark, äh, danach gerichtet. Also die Beamten spielten eine große Rolle und, äh, man hatte so eine gewisse, so ein Gefühl ... Also ich war ja erst fünf Jahre, nicht wahr. Das sind die allerersten Jahre. Ja. Übrigens erinnere ich mich noch an den Untergang der Titanic. Da haben die Leute ... Ja? Ja, nicht. Die war da schon gewesen, aber dass die Leute davon sprachen. Und, äh, das war zwei Jahre vor meiner Geburt, ne? Dass das noch, äh, also in der Erinnerung war. Ja, jetzt wird sie wieder gehoben. Ja, versucht man. Ja. Ähm, also dann diese gesellschaftlichen Verhältnisse, die waren also so, äh, dass man sich wirklich zu ner ~upper class~ gehörend fühlte, als Beamtenkind. Und die anderen, das waren die Straßenkinder. Also ich meine, ich weiß genau von dieser Zeit, wie, wie stark der gesellschaftliche Unterschied war. Ja. Ja. Und mit Straßenkindern spielte man nicht, ne? Außerdem, äh, also, meine Mutter konnte, die konnte noch so eben Kinderwagen führ ..., fahren. Ein Vater, der tat das schon mal gar nicht. Also die Mutter eigentlich auch nicht. Dafür hatte man seine Angestellten. Und was ich so rückblickend, äh, als grotesk empfinde, sie konnte nicht von hier bis da zum, äh, Krämer rüber gehen, ohne einen Hut

Simple Detailed

Custom

Simple search. Terms may be combined with + # -

+ = AND

# = OR

- = NOT

**Keyword**

Search

---

**Annotations**

**Which Annotation Scheme?**

test lusir

work i

family i

war i

UNSET ANNOTATION

• = Document-wide annotation

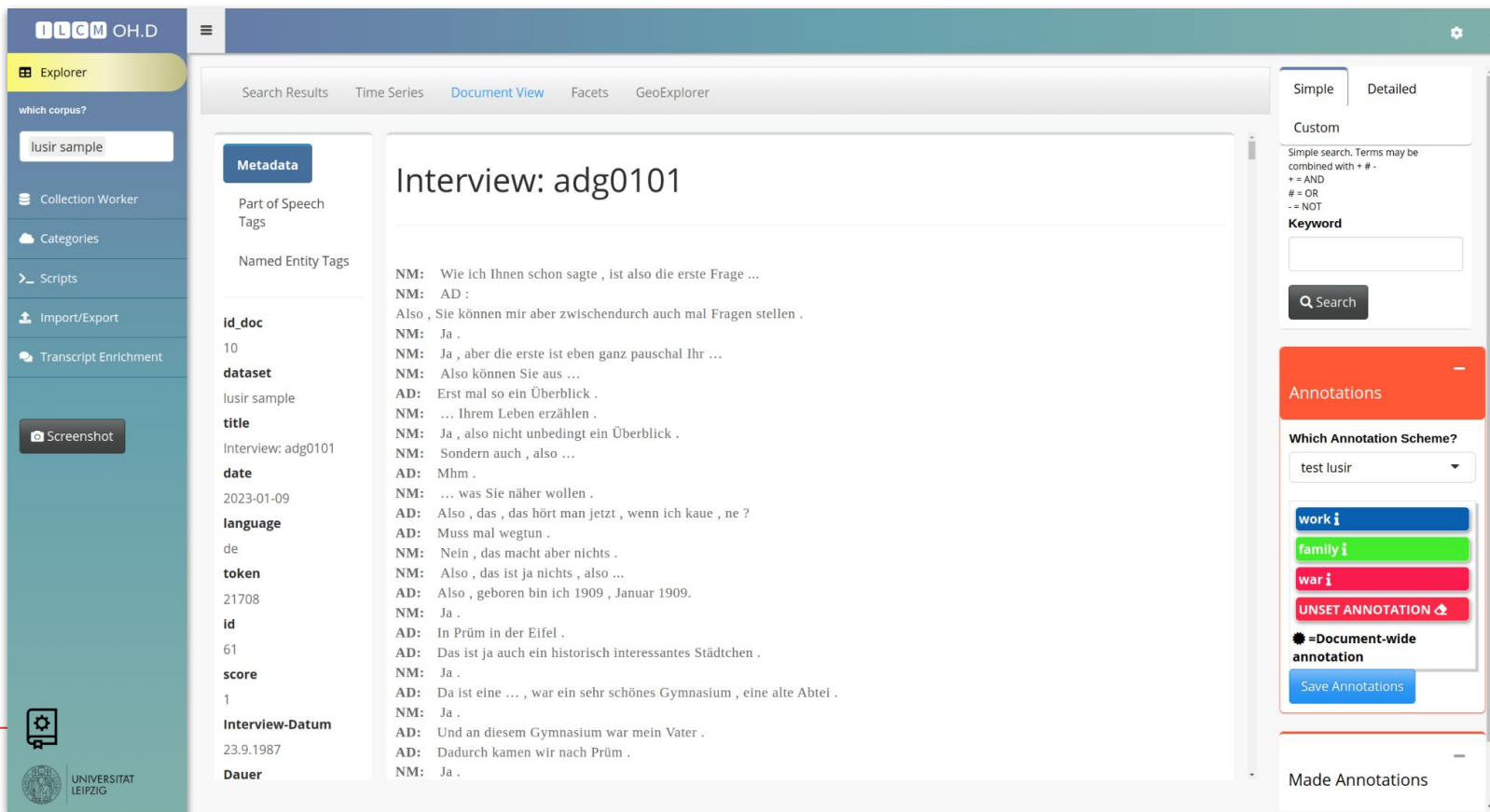
Save Annotations

---

Made Annotations



## Results: Adaptation of document view to interview format



**iLCM OH.D**

Explorer

which corpus?  
lusir sample

Collection Worker

Categories

Scripts

Import/Export

Transcript Enrichment

Screenshot

Search Results Time Series **Document View** Facets GeoExplorer

**Metadata**

Part of Speech Tags

Named Entity Tags

**id\_doc**  
10

**dataset**  
lusir sample

**title**  
Interview: adg0101

**date**  
2023-01-09

**language**  
de

**token**  
21708

**id**  
61

**score**  
1

**Interview-Datum**  
23.9.1987

**Dauer**

**Interview: adg0101**

NM: Wie ich Ihnen schon sagte, ist also die erste Frage ...  
 NM: AD :  
 Also, Sie können mir aber zwischendurch auch mal Fragen stellen .  
 NM: Ja .  
 NM: Ja, aber die erste ist eben ganz pauschal Ihr ...  
 NM: Also können Sie aus ...  
 AD: Erst mal so ein Überblick .  
 NM: ... Ihrem Leben erzählen .  
 NM: Ja, also nicht unbedingt ein Überblick .  
 NM: Sondern auch, also ...  
 AD: Mhm .  
 NM: ... was Sie näher wollen .  
 AD: Also, das, das hört man jetzt, wenn ich kaeu, ne ?  
 AD: Muss mal wegtun .  
 NM: Nein, das macht aber nichts .  
 NM: Also, das ist ja nichts, also ...  
 AD: Also, geboren bin ich 1909, Januar 1909.  
 NM: Ja .  
 AD: In Prüm in der Eifel .  
 AD: Das ist ja auch ein historisch interessantes Städtchen .  
 NM: Ja .  
 AD: Da ist eine ..., war ein sehr schönes Gymnasium, eine alte Abtei .  
 NM: Ja .  
 AD: Und an diesem Gymnasium war mein Vater .  
 AD: Dadurch kamen wir nach Prüm .  
 NM: Ja .

Simple Detailed

Custom

Simple search. Terms may be combined with + # -  
 += AND  
 # = OR  
 -= NOT

**Keyword**

Search

**Annotations**

Which Annotation Scheme?  
test lusir

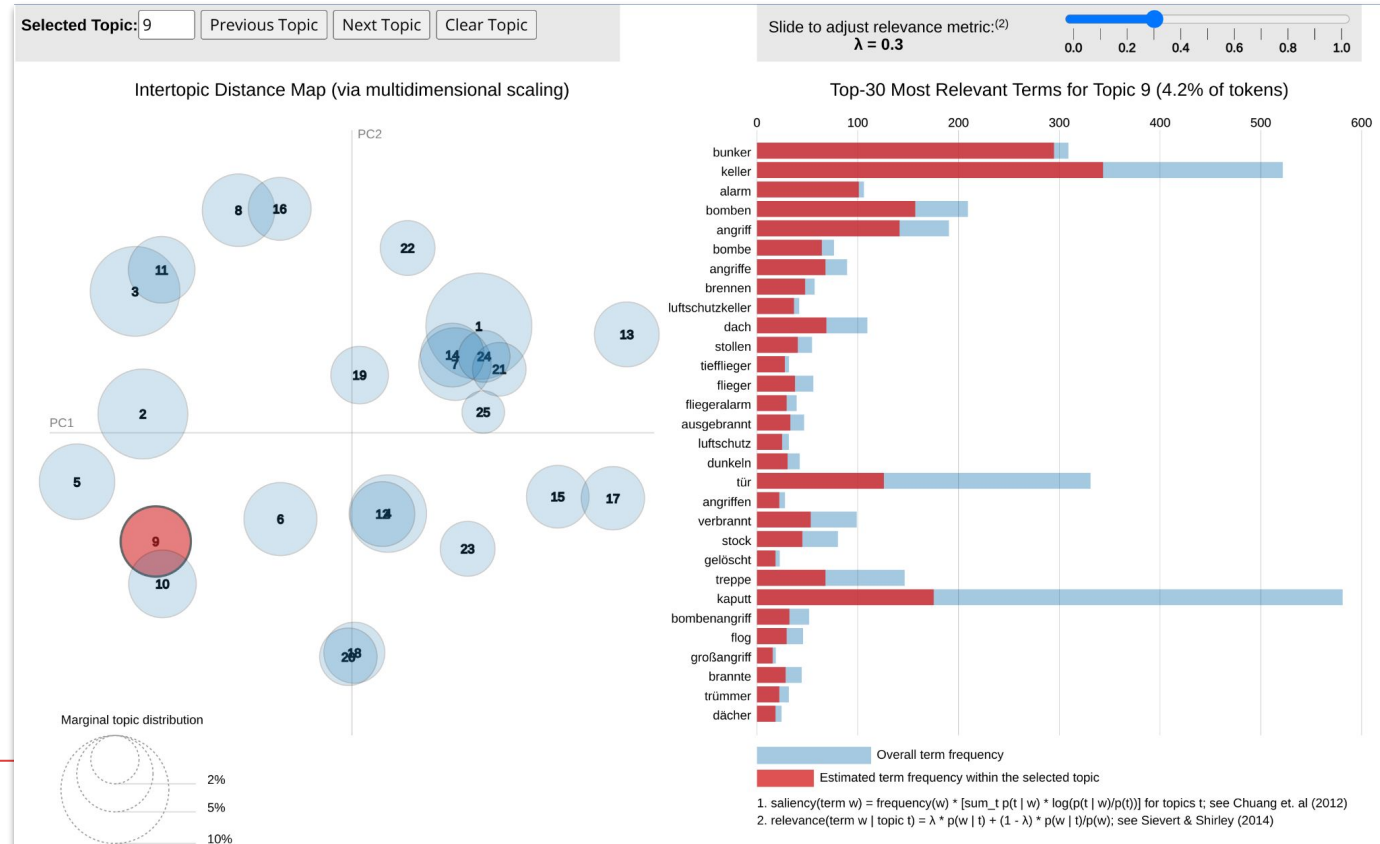
work i  
family i  
war i  
UNSET ANNOTATION

Document-wide annotation  
Save Annotations

Made Annotations

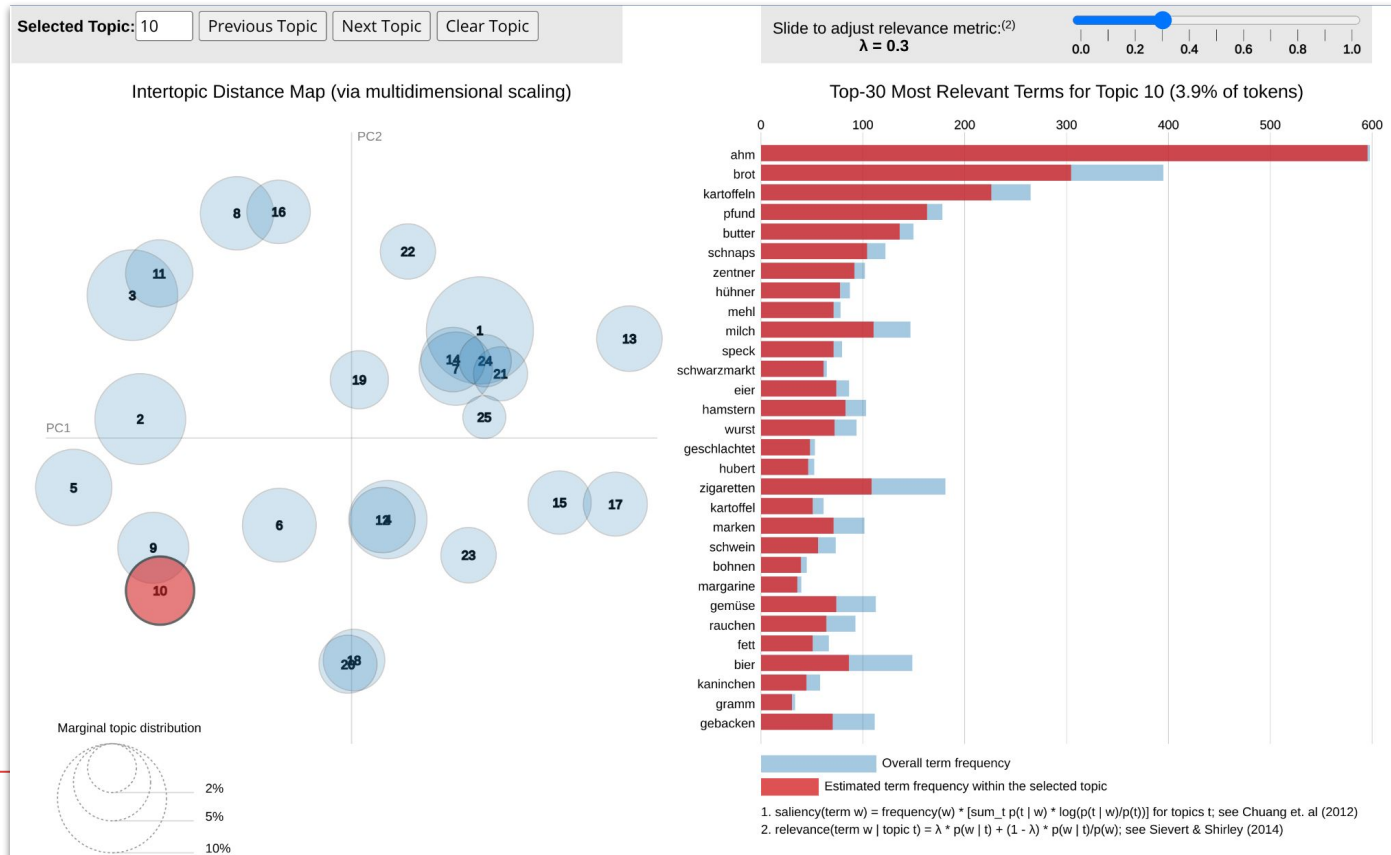
# Results: Topics found

- “Bombings”



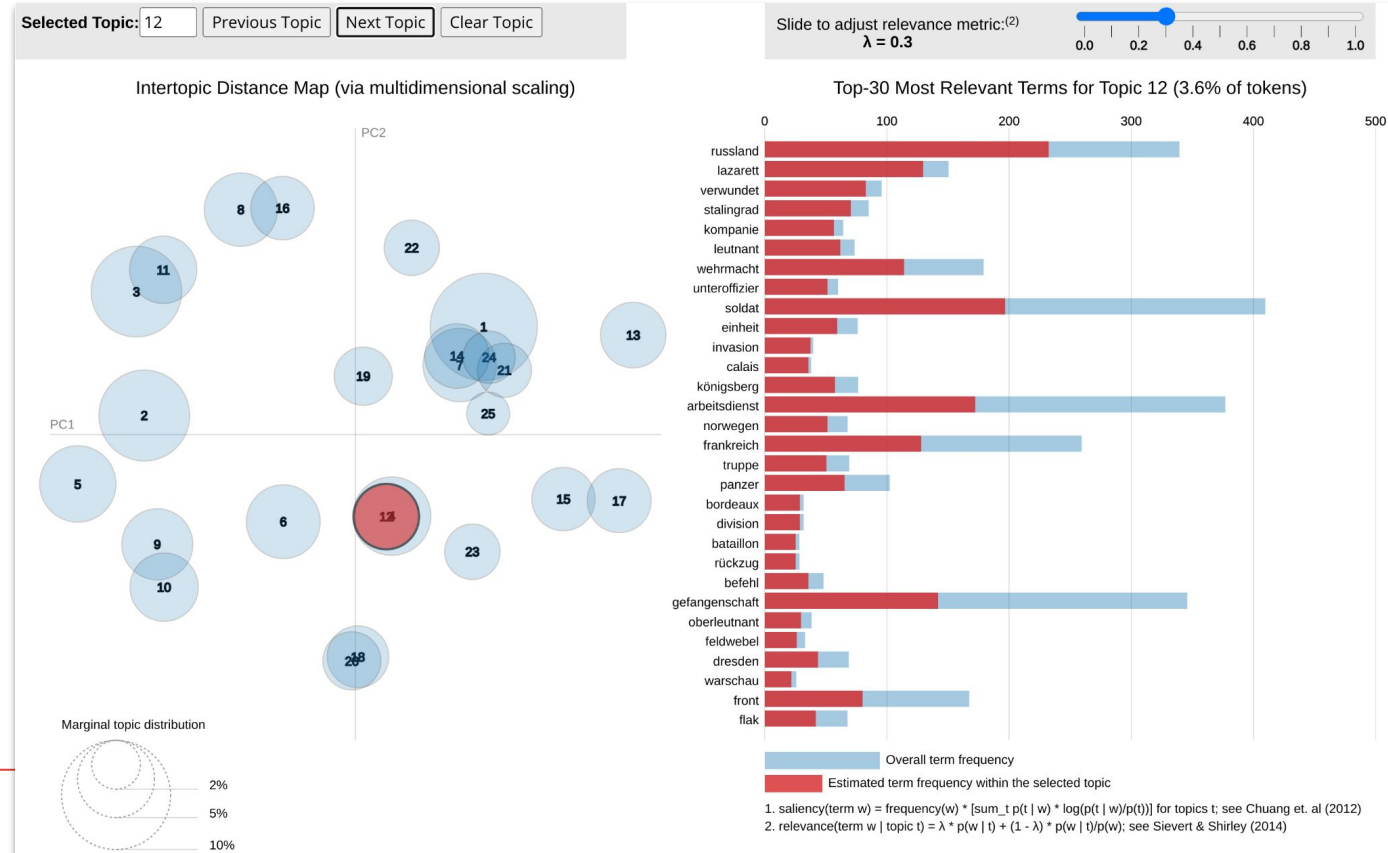
# Results: Topics found

- "Food"



## Results: Topics found

- “Experiences of war”





# Results: Identification of relevant passages

LDA-Vis
Estimated word frequencies
Topic Proportions over time
Date distribution
Coherence
Extract Dictionaries
Detailed Metadata Distribution
Topic Dispersion
Validation
Document Comparison
Document Outlier

Document Clustering
Document Grouping
Model Reproducibility
Topic Topic Connection
Topic Labels

Distribution of topics for chosen document

Topic	Percentage
Topic: 12	27.5%
Topic: 15	14.7%
Topic: 22	10.9%
Topic: 9	7.63%
Topic: 6	7.33%
Topic: 17	6.1%
Topic: 1	5.65%
Topic: 20	3.4%
Topic: 18	3.4%
Topic: 11	3.4%
Topic: 7	3.4%
Topic: 3	3.4%
Topic: 2	3.4%
Topic: 16	3.4%
Topic: 19	3.4%
Topic: 21	3.4%

ich dachte erst mal, äh, bisher haben wir nur ein paar Mal über Termine gesprochen und gestern Morgen habe ich Ihnen dann gesagt, habe ich gedacht: „Na gut, dann haben wir das hinter uns, dann machen wir das also schnell.“ Und gestern Nachmittag habe ich mir eigentlich mal überlegt, was möchte die Dame eigentlich von mir wissen. (...) Das könnte ja eigentlich doch ein bisschen unerfreulich werden, nicht, du hast doch ein bisschen vorzeitig zugesagt. Also, dass die ... ich hätte besser getan, hätte mir zunächst mal ihre Bücher beim (Brockmeier) besorgt und hätte da mal reingeguckt, dann hätte ich es wahrscheinlich gesehen. Also da kann ich Ihnen zusichern, dass, wie gesagt, mit den Sachen kein Missbrauch getrieben wird und ich bin also auch aus persönlichen Gründen daran interessiert, äh, was so über die mittelständischen Unternehmer ... Sie sind ja im Prinzip ein mittelständischer Unternehmer, äh zu erfahren weil ich eben wie gesagt auch vorhabe, äh, darüber eine Doktorarbeit zu schreiben. Also einfach mal zu sehen wie die mittelständischen Unternehmer sich in der Bundesrepublik Deutschland entwickelt haben, weil da gibt es nämlich auch nicht sehr viel auf diesem Gebiet. Also es dreht sich jetzt um Ihre Doktorarbeit. Nein, das verknüpfe ich miteinander. Also die Arbeit in Hagen verknüpfe ich mit meiner Arbeit in Bochum. Das ist ... [lacht]. Man muss ja rationell arbeiten. Fangen wir mal an. Also ich hätte an sich ja ganz gerne Ihren Fragenkatalog vorher mal gerne ... Da habe ich nicht dran gedacht, ich hätte sagen müssen ... Schicken Sie mir Ihren Fragenkatalog ...“. Äh, das machen wir grundsätzlich nicht, weil dann kommt ja nicht, also ... Ja, dann müssen Sie damit rechnen, dass grundsätzlich der Gesprächspartner sagt: „Ja, so haben wir nicht gewettet.“ ne. Hm. Aber ich meine das Prinzip ist ... Denn dass wir schwierige Jahre hinter uns haben, das können Sie sich ja vorstellen. Das ... ich meine, ich bin zwar noch relativ jung, aber das Verständnis für die Entwicklung Ihrer Generation kann ich durchaus aufbringen, dafür bin ich ja Historikerin ... Naja, als der Krieg zu Ende war ich in Ihrem Alter, war ich ... ja ungefähr in Ihrem Alter. Und war **Abiturient** mit Lebenserfahrung, nicht. Weiter nix. Ja. Gehöre zu der Generation die also vor dem Krieg, äh, Abitur gemacht hat und wir mussten ja um einen Studienplatz zu bekommen mussten wir erst mal im **Arbeitsdienst** sein und Wehrdienst absolviert haben. Hm. Nicht, und das habe ich auch gemacht und war also bei Preußens Glorien und als wir dann, weil wirs so gut kannten und der Krieg dann begann, haben sie uns gleich dabehalten. Und dann ging das los und als der Krieg zu Ende war, da war ich zwar **Offizier** und ein bisschen dekoriert auch, nicht, aber mit einigen Blessuren bin ich nach Hause gekommen und war **Abiturient** mit Lebenserfahrung. Hm. Sie sagen Sie sind fast 70. Dann sind Sie 19 ... 18 geboren. Neunzehnhundert. Ich bin nicht fast 70, es fehlen noch ein paar Jahre, noch zwei. Eben. [lacht][lacht] Ich bin 18 geboren. Und haben dann in der Weimarer Republik praktisch, äh, Abitur gemacht. 18 ne 36. 30. Äh, das ist in der NS-Zeit schon. Hm. Vor meinen, vor dem Krieg jedenfalls. Vor dem Krieg. Nicht und war dann im **Arbeitsdienst** und begann den Wehrdienst und dann begann ja, kam ja der Krieg. Hm, äh, nur mal ganz kurz, so zu Information. Sie sind hier in Bochum zur Schule gegangen. Ich bin in Bochum zur Schule gegangen. Hier auf dem ... Habe in Bochum mein Abitur gemacht nicht, und, äh, wie gesagt dann kam der Wehrdienst und als der Krieg zu Ende war, dann war ich zunächst mal 100 Prozent wehrdienstbeschädigt. Und, äh ... Darf ich fragen warum. Haben Sie ... Bitte. Ähm, darf ich fragen ... Ich habe Gott sei Dank in der ganz letzten Phase in Ostpreußen n **Heimatschuss** bekommen. Und zwar bin ich erst im Januar 45 noch nach Ostpreußen gekommen nachdem wir hier diese **Westoffensive**, äh, ja praktisch mehr oder weniger, glücklich oder unglücklich, wie mans will, hinter uns gebracht habe, bin ich als **Offizier** noch wieder raufgekommen, ich sollte nach **Kurland** aber **Kurland** war bereits zu der Zeit schon abgetrennt, das ging nicht mehr. Und habe mit der letzten „Monte Rosa-Fahrt“ den Weg oben nach Pilau gemacht, nach Ostpreußen und wir sind also, äh, nach **Kurland** gar nicht mehr gekommen sondern in Ostpreußen haben wir noch ein bisschen Krieg gespielt. Nicht, und da hab ich dann Gott sei Dank im Februar 45 den **Heimatschuss** bekommen und bin dann im März 45 mit einem der letzten Schiffe, Lazarettsschiffe, bin ich nach Hause gekommen. Also **Heimatschuss** heißt Sie sind **verwundet** worden. Ich bin **verwundet** worden nicht, äh, und äh, bekam dann **Marschbefehl** Richtung Heimat **Lazarett**. Kann man sagen, dass Sie noch Glück gehabt haben. Das war mein Glück. Deswegen sage ich ja **Heimatschuss**. Hm. Der **Heimatschuss** war damals bei den Soldaten sehr beliebt. [lacht] Ja, ehrlich. In der Tat, ist ja klar. Hm, ähm, Sie sagten Sie waren an der **Westoffensive** gewesen, ist das, also meinen Sie damit die **Normandie**. Also eines Einmarsch der ... Ich habe die **Westoffensive**, äh, nachher bis **Bastogne** noch mitgemacht. Also ... Jetzt überlege ich gerade. Das heißt Sie waren in **Frankreich stationiert**. Nein. **Frankreich** habe ich nie gesehen. Leider wird das oder wird das nachher (...). Das ist doch die **Westoffensive**, nicht oder. Oder vertue ich mich jetzt. Ja, das schon, aber ich hab ziemlich, ziemlich viel während des Krieges von der Welt gesehen. Bin erst gleich am 8. April 1940 nach **Norwegen** rübergegangen und bin dann bis 43 in **Norwegen** gewesen. Was waren Sie denn. Ich war Panzerjäger. Und, äh, habe dann nachher, gehörte nachher zu einer Panzerabteilung, Panzerjägerabteilung und dann wurde es ja im Westen ziemlich **brenzlich** und im Osten noch viel **brenzlich**. Hm. Und dann wurden wir da rausgelost, wir kamen dann zum Aufpäppeln so n bisschen, nach, äh, **Danemark**. Haben uns in Buttersopf gesetzt so ungefähr und dann gings ab in Richtung, zunächst mal übers Heimatland, über die Heimat-Garrison, äh, nach, äh, in Richtung Osten. Und zwar im Weichselbogen. Hm. Und im Weichselbogen, äh, da wurden wir dann **ausgeladen**, wir sollten dann irgendwie aufgeteilt werden, aber das war so ein (...). Transport, da war der **Russe** inzwischen **durchgebrochen** bei **Radom**, war der **Russe** **durchgebrochen** und dann wurde ganz schnell noch eine **Kompanie** **zusammengestellt** und da war ich der älteste, dienstälteste **Offizier** der da war. Wurde ich also der **Kompanieführer** und ich kriegte also die ganzen Leute die da waren zu ... an den Hals geworfen und mussten oben eine, eine Höhe besetzen. Und warteten auf den Russen. Und an diesem, am ersten Abend, das vergesse ich in meinem Leben nicht, da wurde ich bereits von den Russen drüben mit Lautsprechern **empfangen** ... Mit Lautsprechern. Damals, ich war damals junger **Leutnant**. Herr **Leutnant Wiemann** wir begrüßen Sie. Wir freuen uns, dass Sie hier sind. Wir werden Sie heute Nacht besuchen, haben Sie keine Angst, wir holen Sie, es geschieht Ihnen gar nicht, wenn Sie keine Gewissheit (...). Und an der Nacht kamen die. Was. In der Nacht kamen die. Das ist doch

Most relevant words for chosen topic



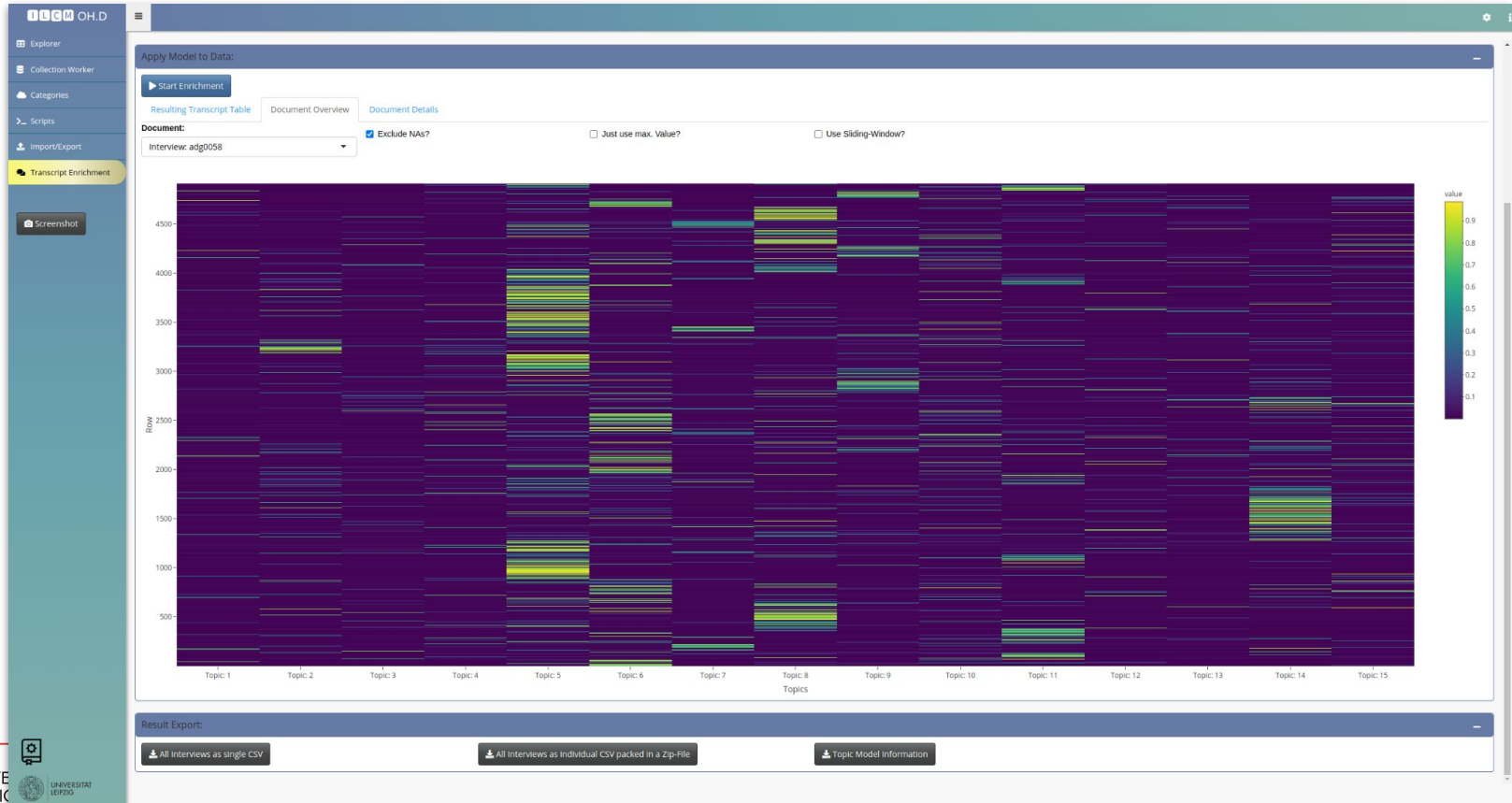


## **Results:** Transcript Enrichment with topic information

- Transcript Enrichment:
  - Application of the overall topic model found on all texts to individual chunks of individual interviews
  - chunks =  $n$  speech acts
  - Each chunk is assigned a probability distribution over the set of topics
  - Visualization of the distribution allows identification of individual sections that deal with relevant topics



# Results: Transcript enrichment with topic information



## Livedemo

<https://berdilcm.ddnss.de/>



Oral-History.Digital



UNIVERSITÄT  
LEIPZIG

**Thank you**

**Questions?**

**Christian Kahmann**

[kahmann@informatik.uni-leipzig.de](mailto:kahmann@informatik.uni-leipzig.de)