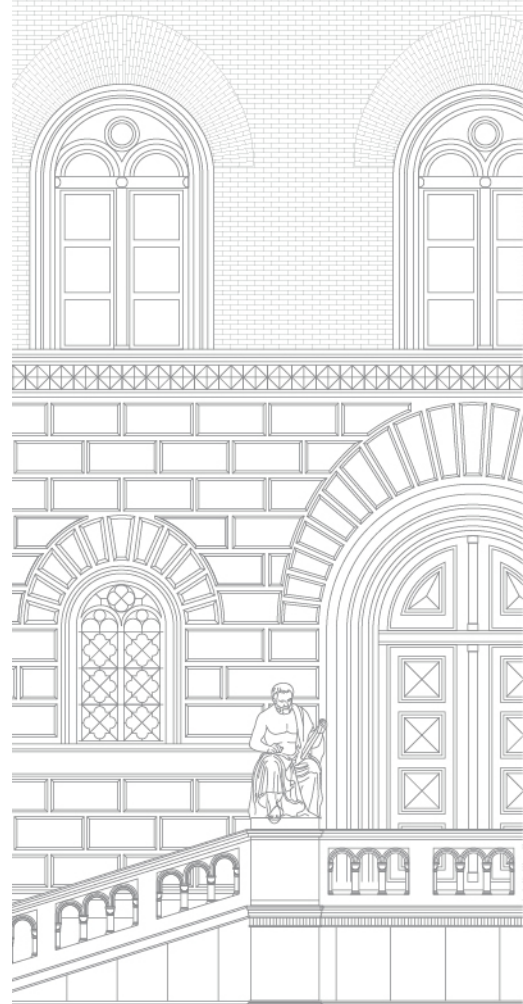


## Webarchive und Digital Humanities

DFG-Projekt der Bayerischen  
Staatsbibliothek in Kooperation mit der  
Universität Passau



# Herausforderungen

Datenvielfalt

Datenmenge

Datenbereitstellung

Interdisziplinär

wissenschaftliche Anforderungen  
— technische/organisatorische  
Rahmenbedingungen

Aufwand für selektive  
Webarchivierung

# Volltextsuche mit Named Entity Recognition (NER)

Grenze Search

---

**DOMAIN**

<input type="checkbox"/> die-linke.de	32,829
<input type="checkbox"/> fdp.de	25,439
<input type="checkbox"/> skakeller.de	23,511
<input type="checkbox"/> afd.de	18,115
<input type="checkbox"/> cdu.de	11,818
<a href="#">+ More</a>	

**ORT**

<input type="checkbox"/> Europa	79,809
<input type="checkbox"/> Berlin	71,817
<input type="checkbox"/> Deutschland	58,316
<input type="checkbox"/> europäischen	50,082
<input type="checkbox"/> Brüssel	39,993
<a href="#">+ More</a>	

**PERSON**

<input type="checkbox"/> Rue Wiertz	23,176
<input type="checkbox"/> Vita	23,035
<input type="checkbox"/> Robert Schuman	22,951
<input type="checkbox"/> Bundesschiedsgericht	17,921
<input type="checkbox"/> Mo	17,365

**FAQ: Flucht und Migration | Christlich Demokratische Union Deutschlands**

**URL:** <https://www.cdu.de/faq-asyl>

**Textauszug:** ... und Migration FAQ: Migration ordnen, steuern, **begrenzen** Zur aktuellen Debatte erreichen uns viele  
Zuschritten...

**Gecrawlt am:**  
28.3.2019  
3.4.2019

**Ohne Titel**

**URL:** [https://www.skakeller.de/fileadmin/material/Migration/pdf/SkaKeller\\_Briefing\\_EuropaeischerGrenzundKu%CC%88stenschutz.pdf](https://www.skakeller.de/fileadmin/material/Migration/pdf/SkaKeller_Briefing_EuropaeischerGrenzundKu%CC%88stenschutz.pdf)

**Textauszug:** ...1 Brüssel, den 01. Juli 2015 EUROPÄISCHER **GRENZ-** UND **KÜSTENSCHUTZ** Mehr Macht für  
Frontex...

**Gecrawlt am:**  
24.3.2019  
31.3.2019

# Volltextsuche mit Named Entity Recognition (NER)

## Herausforderungen:

- Genauigkeit
- Boilerplate-Inhalte
- Disambiguierung

ORT	
<input type="checkbox"/> Berlin	190,803
<input type="checkbox"/> Alexanderstraße	186,550
<input type="checkbox"/> Europäische	80,748
<input type="checkbox"/> Deutschland	63,178
<input type="checkbox"/> Europa	49,639
<a href="#">+ More</a>	
PERSON	
<input type="checkbox"/> Katja Kipping	39,446
<input type="checkbox"/> Bernd Riexinger	32,722
<input type="checkbox"/> Dietmar Bartsch	15,957
<input type="checkbox"/> Merkel	14,351
<input type="checkbox"/> Angela Merkel	12,451
<a href="#">+ More</a>	

# Sammlungsausbau: Suche nach relevanten Websites

Volltextsuche und  
Textähnlichkeitssuche

## Herausforderungen:

- Expert\*innenwissen formalisieren
- Nachvollziehbarkeit
- Formulierung von Suchanfragen

Filter nach Schlagwort: Alte Synagoge Binswangen (Binswangen)  
Archiv  
Archivalien  
Augsburg, Anita, 1857-1943  
Bayern

Suche nach: Vielfalt der bayerischen Sprache

Suche starten

▼ 1. **bund-bairische-sprache.de**

Link 1

URL bund-bairische-sprache.de

Kontext Bund Bairische Sprache

▶ 2. phil.uni-passau.de

▶ 3. philhist.uni-augsburg.de

▶ 4. bayrisch-lernen.de