

# Erschließung eines Netzwerks von Named Entities aus Nachrichtenartikeln am Beispiel von [www.tagesschau.de](http://www.tagesschau.de)

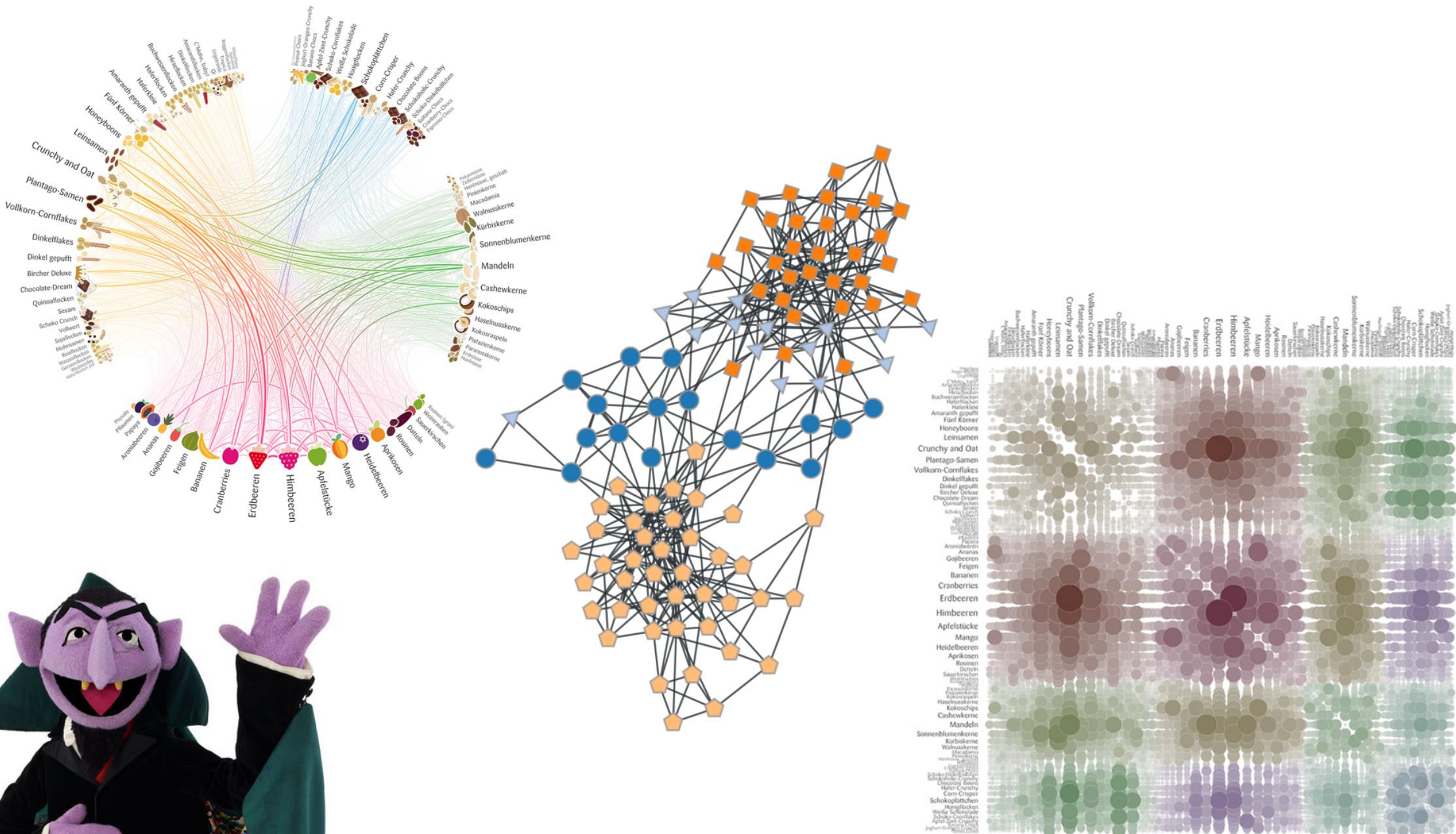
*Analyse semi- und unstrukturierter Daten – MBB WS22 (K)*

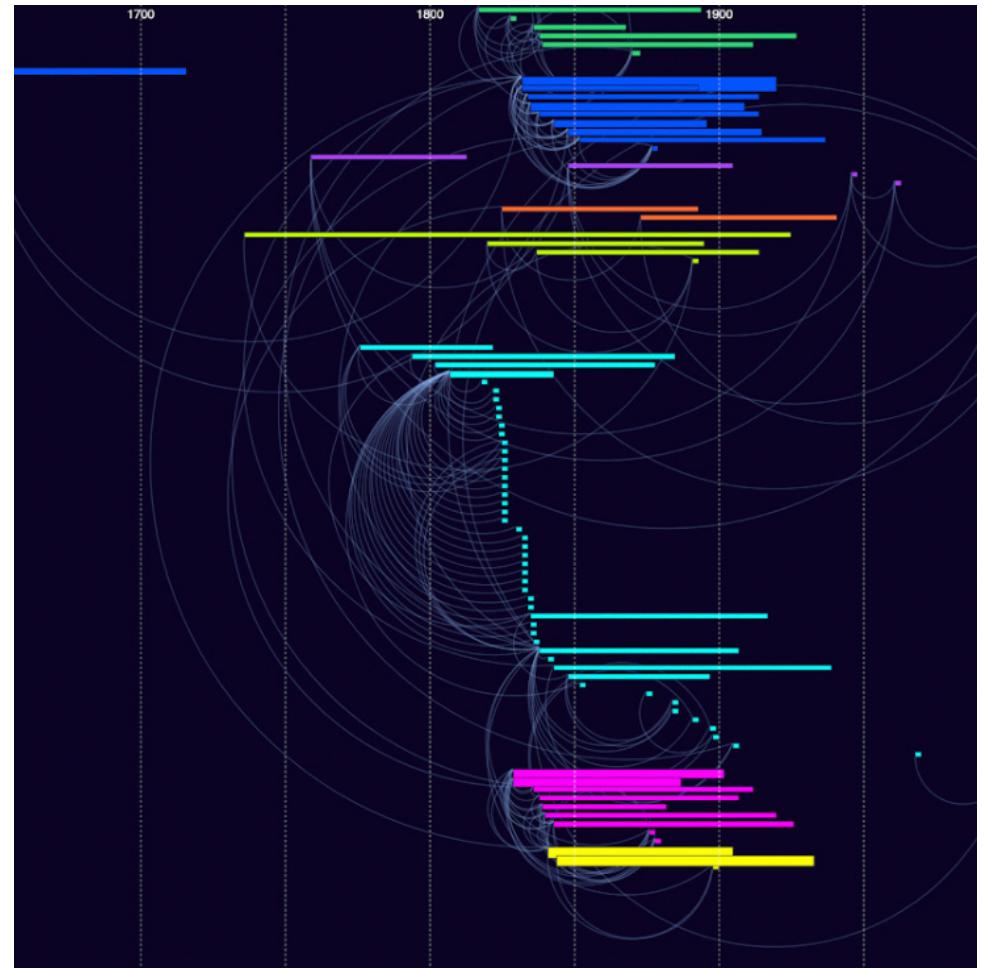
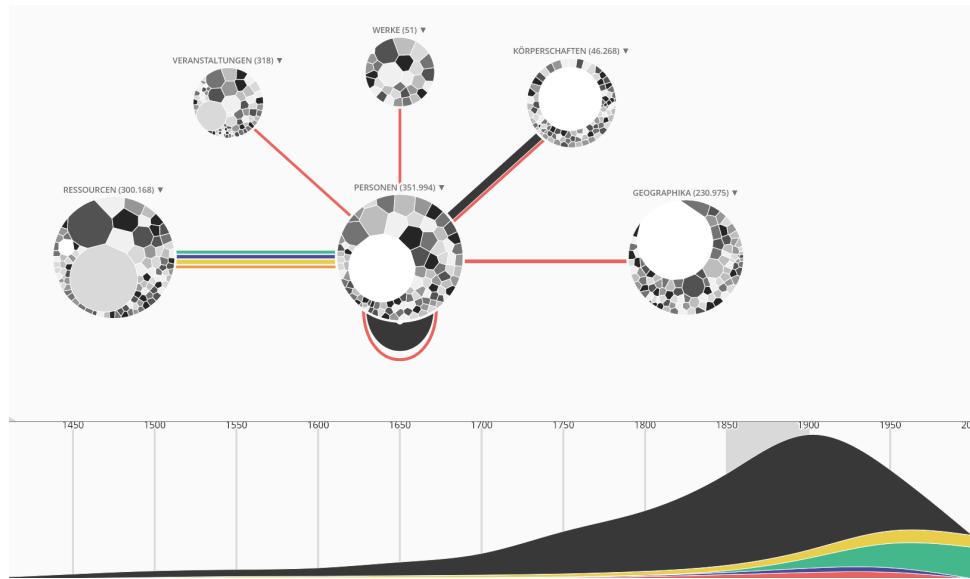
Patrick Haas – 311431

*p@trickhaas.de*

Mittwoch, 12. Juli 2023

# Motivation und Inspiration





# SoNAR (IDH): Interfaces to Data for Social Historical Network Analysis and Research

Revision abgelehnt

## BGH bestätigt Haftstrafe gegen Starkoch Schuhbeck

Der Starkoch Schuhbeck muss wohl ins Gefängnis. Das Landgericht München I hatte ihn wegen Steuerhinterziehung zu einer Haftstrafe verurteilt - das hat der BGH weitgehend als rechtskräftig bestätigt.

| mehr

19.06.2023 • 17:02 Uhr



Hohe Summe und umstrittene Ziele

## Top-Ökonomen kritisieren Intel-Subvention scharf

Mit einem milliardenschweren Nachschlag bei den Subventionen hat der Bund die Hängepartei um die Chipfabrik von Intel in Magdeburg beendet. Ökonomen sehen die hohe Fördersumme und das Vorgehen kritisch. | mehr

19.06.2023 • 16:57 Uhr



Chinas Ministerpräsident in Berlin

## "Die Tradition der Freundschaft hochhalten"

Chinas neuer Ministerpräsident Li hat zum Auftakt seiner Deutschland-Visite dafür geworben, die Zusammenarbeit beider Länder zu vertiefen und Differenzen zu überwinden. Zu den Themen dürfte neben dem Klimawandel auch der Ukraine-Krieg zählen. | mehr

19.06.2023 • 16:34 Uhr



KOMMENTAR Neuausrichtung der CDU

## Was soll das für eine Partei werden?

Die CDU ringt um Positionierung und Neuaufstellung. Dabei wirft nicht nur die heftig kritisierte Rede von Eisschnellläuferin Pechstein in Polizeiuniform die Frage auf: Welche Partei will die CDU sein? Von Sabine Henkel. | mehr

19.06.2023 • 16:32 Uhr



Streit um EU-Naturschutzrecht

## Umweltbündnis gibt Lemke "Rückenwind"

Mit mehr als 800.000 Unterschriften appelliert die Initiative "Restore Nature" an Umweltministerin Lemke, das geplante EU-Naturschutzgesetz gegen konservative Kräfte zu verteidigen. Die kritisieren das Gesetz als "Verbotspolitik". | mehr

19.06.2023 • 16:06 Uhr

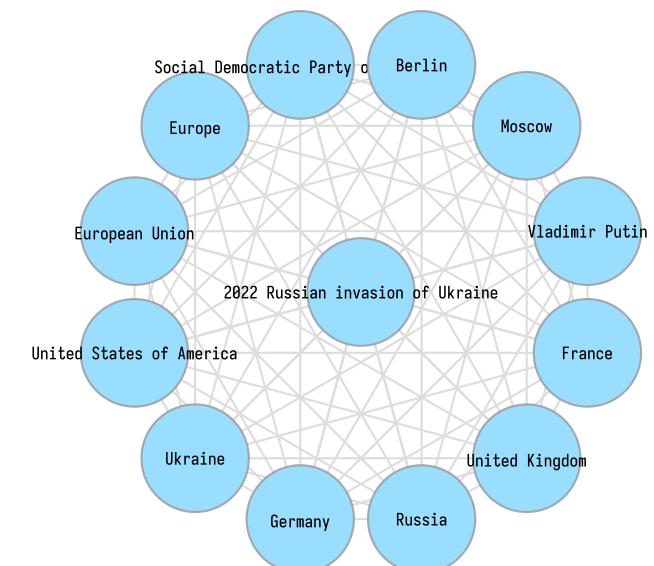


HINTERGRUND Deutsche Autoindustrie

## Ungewisse Zukunft in der postfossilen Welt

Die Autobranche steht vor dem größten Wandel ihrer Geschichte. Kritiker sagen, dass viele deutsche Firmen zu spät dran sein könnten auf dem Weg in die Elektromobilität. Andere haben die Zeichen der Zeit verstanden. Von Tim Diekmann. | mehr

19.06.2023 • 15:55 Uhr



# Konzept

»[...] Die Verurteilung von  
Starkoch Alfons Schuhbeck wegen  
Steuerhinterziehung ist nach einem  
Urteil des Bundesgerichtshofs  
weitgehend rechtskräftig. Im Oktober  
2022 hatte das  
Landgericht München I Schuhbeck  
zu drei Jahren und zwei Monaten  
Gefängnis verurteilt [...].«

Named Entity Recognition

»[...] Die Verurteilung von  
Starkoch Alfons Schuhbeck wegen  
Steuerhinterziehung ist nach einem  
Urteil des Bundesgerichtshofs  
weitgehend rechtskräftig. Im Oktober  
2022 hatte das  
Landgericht München I Schuhbeck  
zu drei Jahren und zwei Monaten  
Gefängnis verurteilt [...].«

Named Entity Disambiguation

»[...] Die Verurteilung von Starkoch Alfons Schuhbeck wegen Steuerhinterziehung ist nach einem Urteil des Bundesgerichtshofs weitgehend rechtskräftig. Im Oktober 2022 hatte das Landgericht München I Schuhbeck zu drei Jahren und zwei Monaten Gefängnis verurteilt [...].«

»[...] Drei Jahre und zwei Monate Haft ohne Bewährung wegen Steuerhinterziehung - so lautete in der vergangenen Woche das Urteil gegen den bekannten Fernsehkoch Alfons Schuhbeck. Schuhbeck will nun Revision einlegen [...].«

**Entity Linking** anhand einer Wissensbasis – hier per *Wikidata* (Wikifizierung).

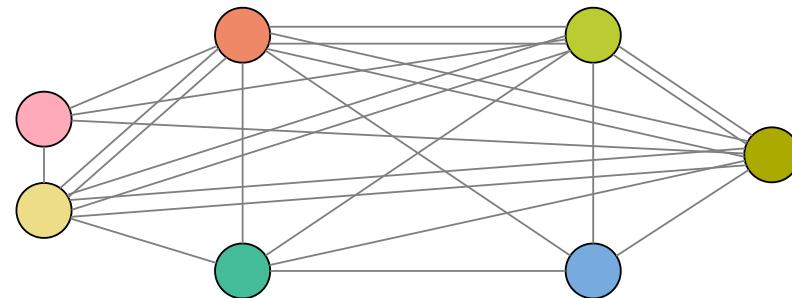
»[...] Die Verurteilung von Q45443 wegen Q17354131 ist nach einem Q3769186 des Q687323 weitgehend rechtskräftig. Im Oktober 2022 hatte das Q1802882 Q45443 zu drei Jahren und zwei Monaten Q841236 verurteilt [...].«

»[...] Drei Jahre und zwei Monate Q841236 ohne Bewährung wegen Q17354131 - so lautete in der vergangenen Woche das Q3769186 gegen den bekannten Q45443. Q45443 will nun Q1523791 einlegen [...].«

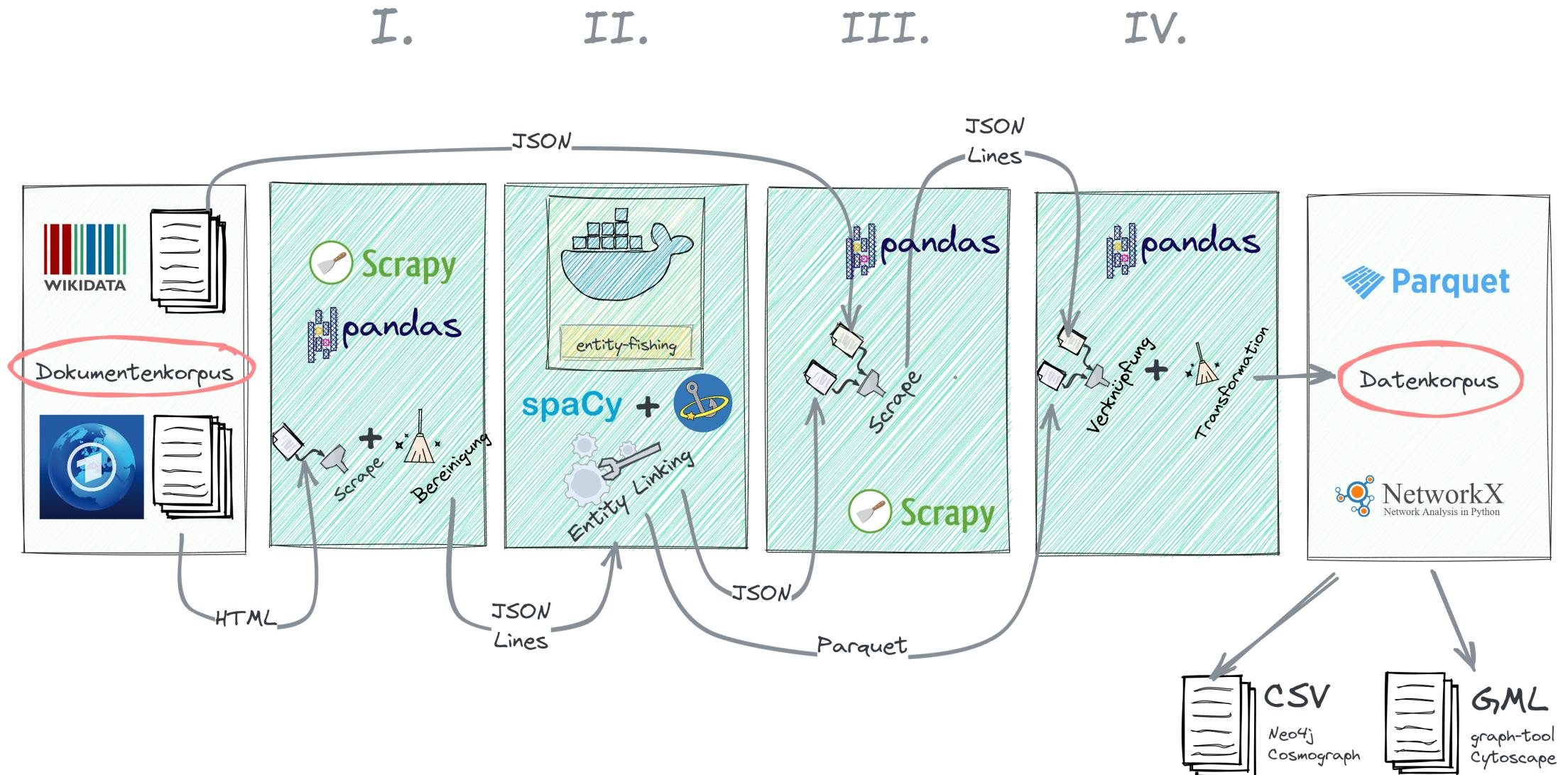
**Entity Recognition + Entity Disambiguation + Referenzierung in eine Wissensbasis = End-to-end Entity Linking**

»[...] Die Verurteilung von Q45443 wegen Q17354131 ist nach einem Q3769186 des Q687323 weitgehend rechtskräftig. Im Oktober 2022 hatte das Q1802882 Q45443 zu drei Jahren und zwei Monaten Q841236 verurteilt [...].«

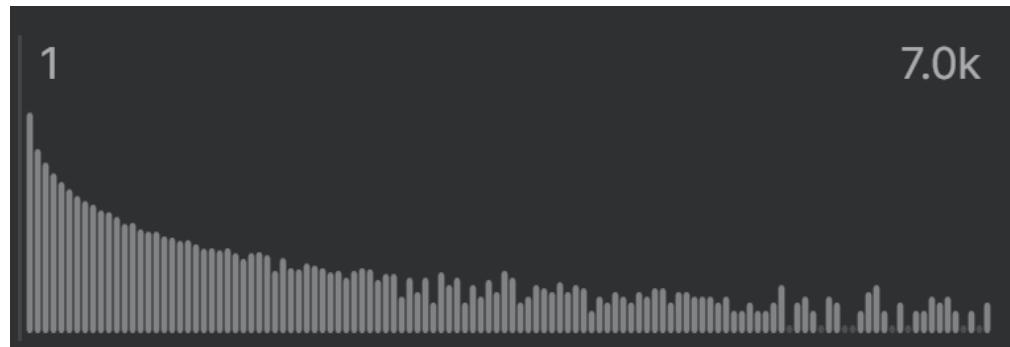
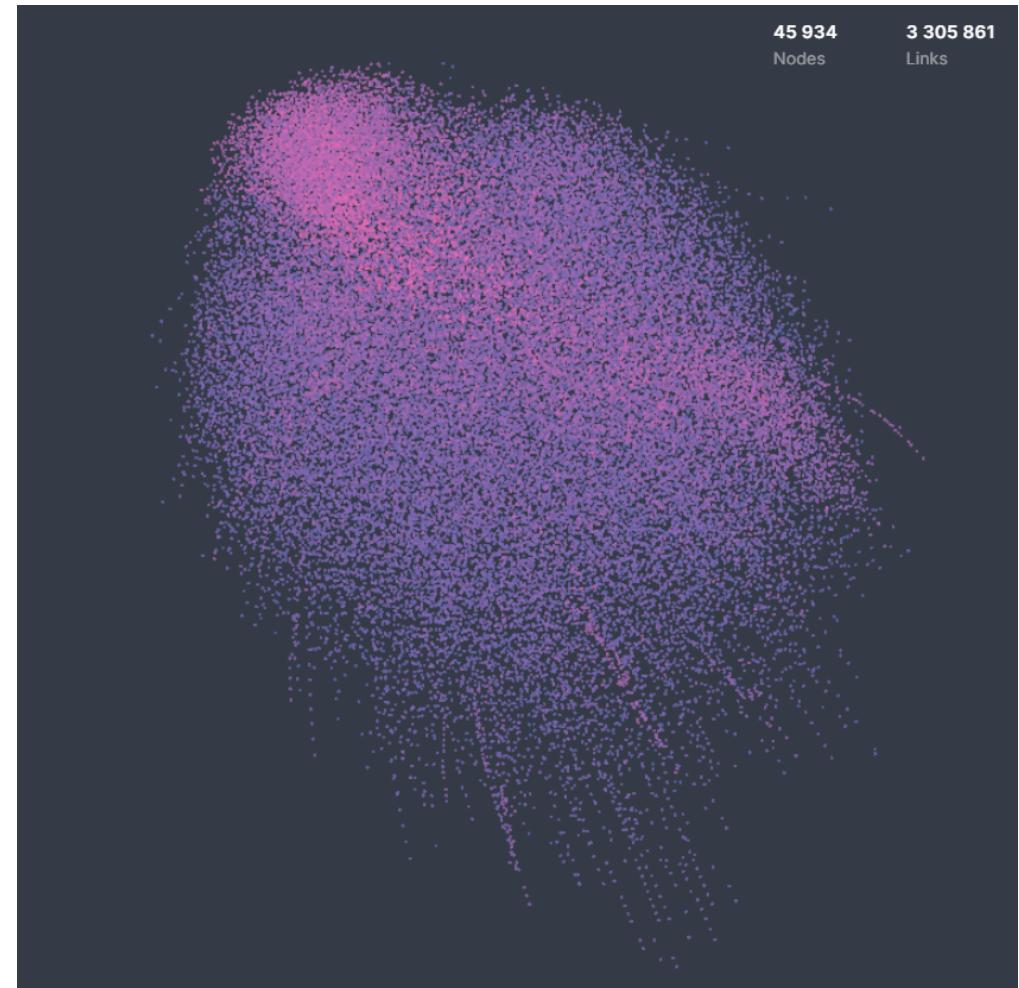
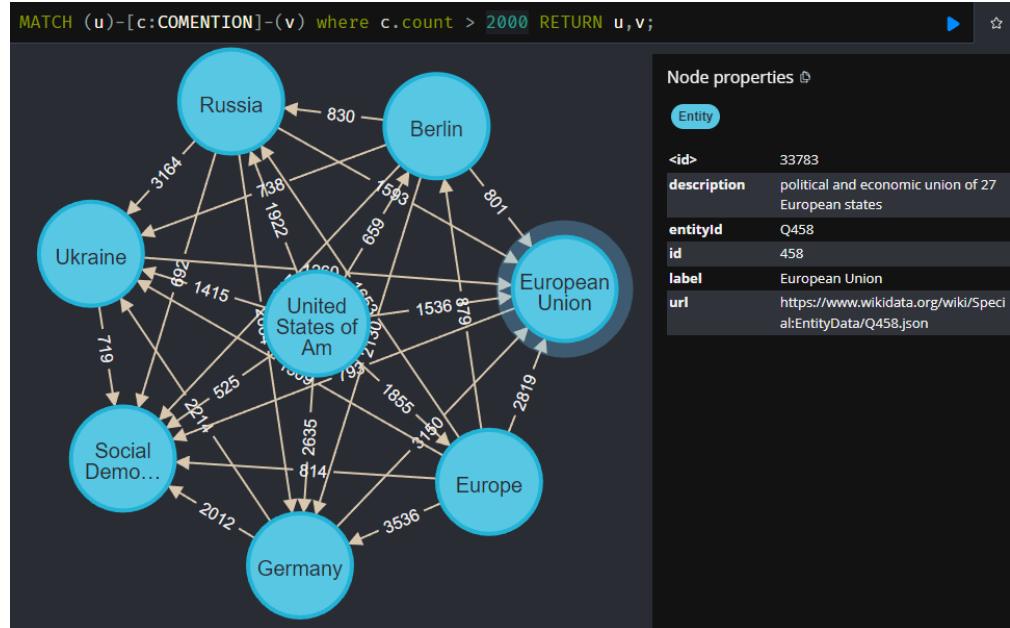
»[...] Drei Jahre und zwei Monate Q841236 ohne Bewährung wegen Q17354131 - so lautete in der vergangenen Woche das Q3769186 gegen den bekannten Q45443. Q45443 will nun Q1523791 einlegen [...].«



# Daten-Pipeline



# Technischer Durchstich



# Herausforderungen

# Entity Linking

1. Named Entity Recognition
2. Rückgriff auf eine Wissensbasis
3. Kandidatenauswahl
4. Bewertung anhand des lokalen Kontexts



Starker Bezug zur gesprochenen Sprache

Deutsche Modelle existieren, aber die aktuelle Forschung konzentriert sich auf das Englische

# Betrachtete Optionen

*REL: Radboud Entity Linker*

Kein deutsches Modell

Trainierbar, aber: deutsche  
Trainings-/Test-/Validierungsdaten  
erforderlich

Daher für diese Arbeit **verworfen**

*entity-fishing*

Deutsches Modell

Mit Standardinstallation niedriger  
Durchsatz – ca. 40 Artikel/min.



**10 Stunden für 25.000 Artikel**

# Durchsatzoptimierung

Lokale Installation des *entity-fishing*-Webservice

Deployment als Docker-Container

Wissensbasis + englisches +  
deutsches Modell = ~60 GiB

Entity Linking ist unabhängig von  
anderen Artikeln

Wissenbasis und Modelle passen  
vollständig ins RAM der Workstation

Gut parallelisierbares Problem!

 33 Minuten für 25.000 Artikel

# Ausblick

**Netzwerkanalyse:** Gradverteilung – Assortativität – Zentralität – usf.

**Vergleich von Netzwerken:** Weitere Nachrichtenportale – deutsch, englisch, französisch, ...

**Quantifizierung der EL-Fehlerquote:** Auf dem Realdatensatz schwer bewertbar – ggfs. durch Vergleich mehrerer EL-Modelle

**Besseres Entity Linking –** deutsches Training/Finetuning anderer EL-Systeme:

- Radboud Entity Linker
- SBB-NED
- REBEL
- GENRE/mGENRE

Vielen Dank! 🌙