

Integration quantitativer und qualitativer Textanalyse

Erschließung und Analyse großer Textkorpora mit Hilfe
maschinellen Lernens am Beispiel der Migrationsdebatte in
Deutschland im Jahr 2015

Frühjahrstagung 2022 des AK Methoden der DEval
30.6.–1.7.2022

Sebastian Haunss¹

André Blessing², Nico Blokker¹, Erenay Dayanik², Jonas Kuhn², Gabriella Lapesa², Sebastian Padó²

¹ SOCIUM, University of Bremen, ² IMS, University of Stuttgart

Forschungsfrage

- Welche Akteurskoalitionen bestimmten im Jahr 2015 den Migrationsdiskurs in Deutschland

Daten

- Zeitungsartikel

Methode

- Diskursnetzwerkanalyse

Forschungsfrage

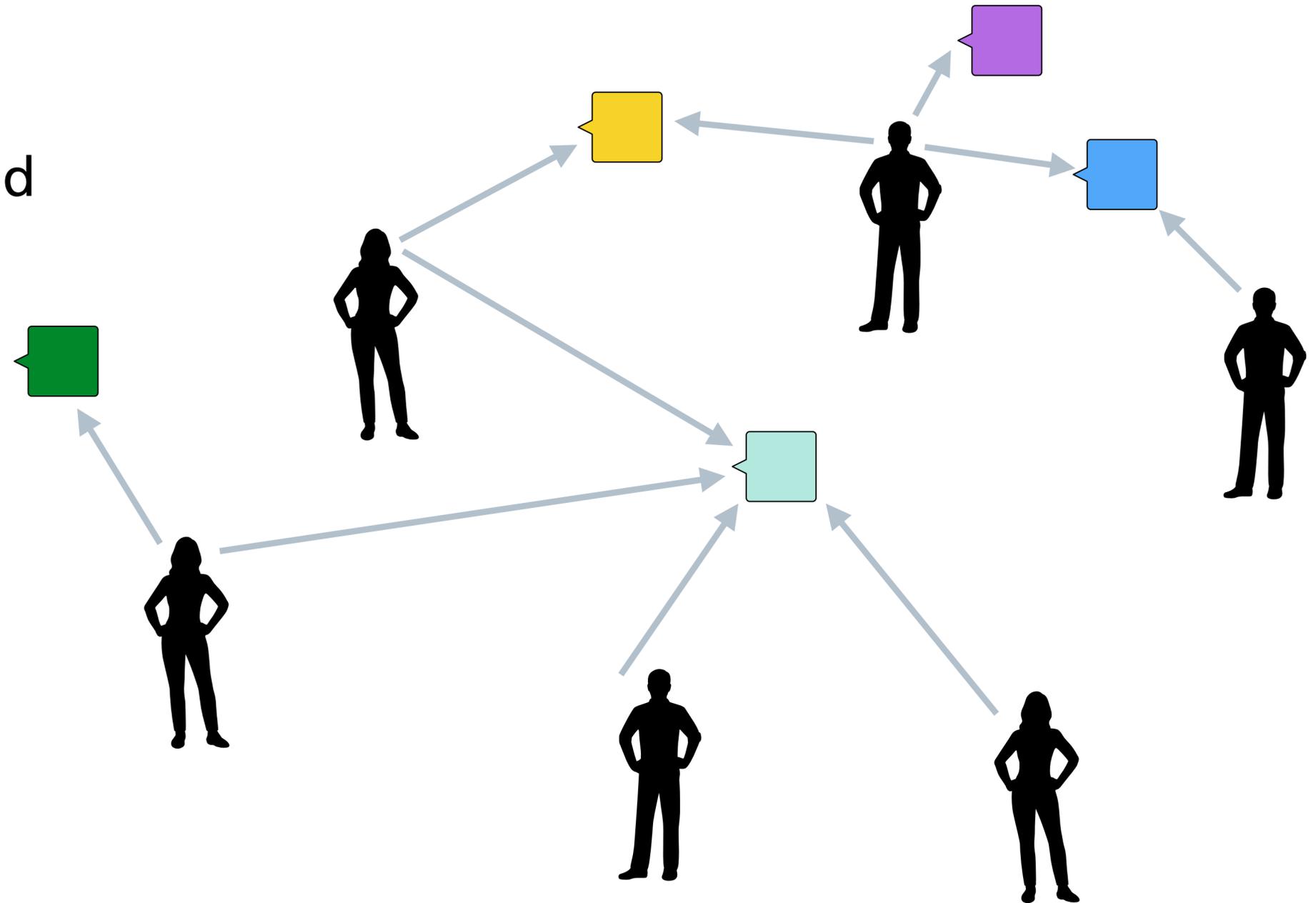
- Welche Akteurskoalitionen bestimmten im Jahr 2015 den Migrationsdiskurs in Deutschland

Daten

- Zeitungsartikel

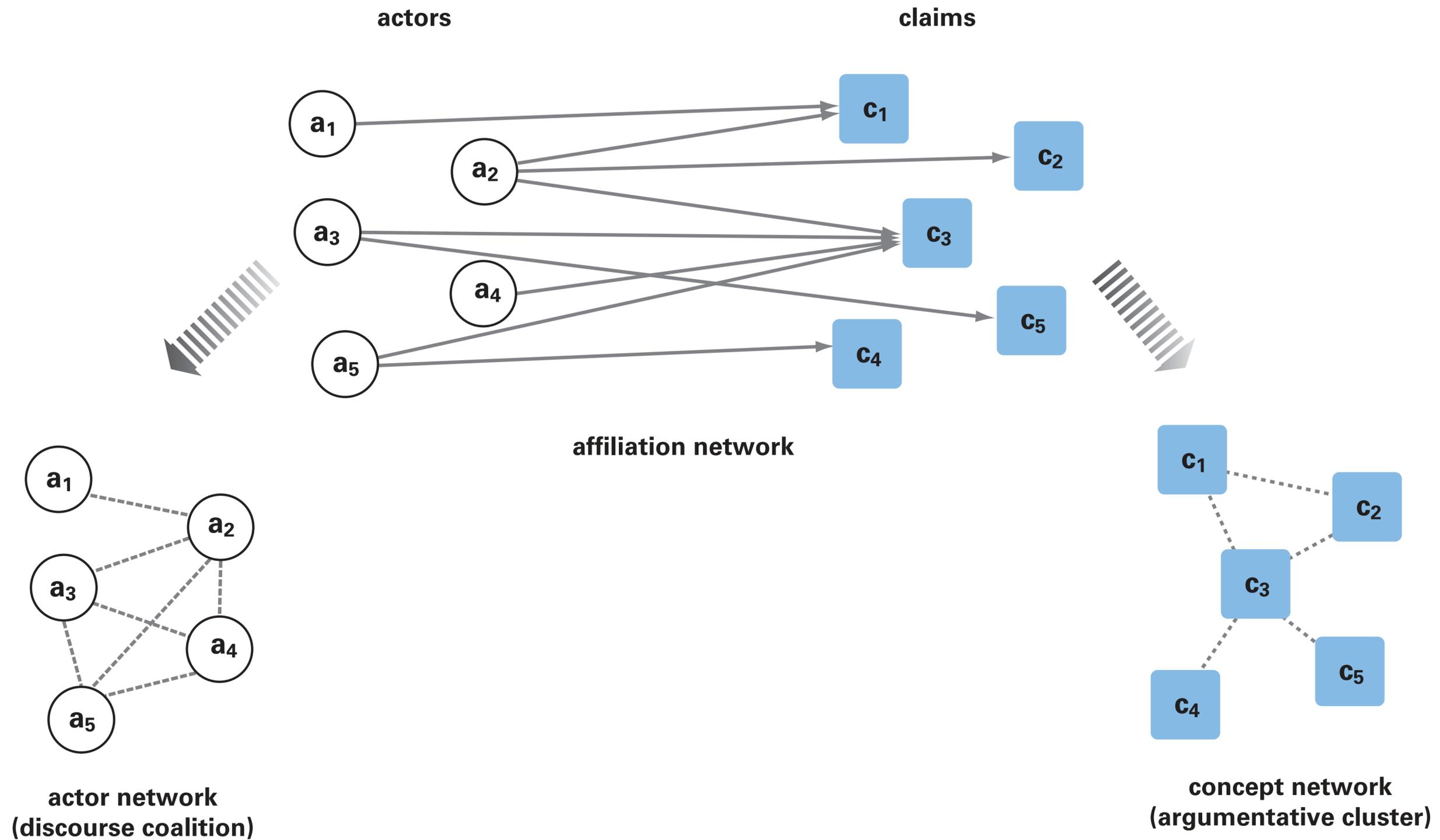
Methode

- Diskursnetzwerkanalyse



Indirekte Kommunikation

Diskursnetzwerke



Das Problem

- Workflow
 - ➔ Text-/Artikelauswahl
 - ➔ Claim Annotation
 - ➔ Sprecher-Identifizierung
 - ➔ Claim Klassifizierung

Restkodierbuch NZZ		SYST	LO	DR	LP	ML	SP	SP	SYST	LO	DR	LP	ML	SP	SP	
2	1	Neue Zürcher Zeitung	D	LO												
		Demonstrationen gegen das Welthandelssystem	7	11	1700	757	9	26	1	700	757	9				
	3	Von Wartenweiler, R.														
	4	840 Wörter														
	5	18 Mai 1998														
	6	Neue Zürcher Zeitung														
	7	NEUZZ														
	8	Deutsch														
	9	Besuchen Sie die Website der führenden Schweizer Internationalen Tageszeitung unter http://www.nzz.ch														
	10	Vor der Eröffnung des WTO-Ministertreffens in Genf														
	11	Das multilaterale Welthandelssystem feiert an der diesjährigen WTO-Ministertagung in Genf sein 50-Jahr-Jubiläum. Die Sitzungen im Völkerbundpalast sind entsprechend den Zufälligkeiten der Präsenz der verschiedenen Persönlichkeiten und zulasten des festlichen Charakters kunterbunt zusammengestellt. Das frühzeitig aufgebaute Sicherheitssystem ist bereits im Vorfeld der Zusammenkunft von Demonstrationen gegen die Globalisierung und gegen den liberalen Welthandel getestet worden.	D	LO												
4			8	26	1	700	757	9	8	26	1	700	757	9		
	12	Ww. Genf, 17. Mai														
	13	Im Zuge des Zusammenbruchs des Systems der kommunistischen Planwirtschaft und des Falls der Berliner Mauer sind die neunziger Jahre mit einer Euphorie über die Geltungskraft marktwirtschaftlicher Prinzipien «eingeläutet» worden. Von diesem Elan profitierte auch die Uruguay-Runde, mit deren Abschluss letztlich das 1948 geschaffene Provisorium des Allgemeinen Zoll- und Handelsabkommens (Gatt) am 1. Januar 1995 durch die auf Dauer angelegte Institution der Welthandelsorganisation (WTO) abgelöst wurde. Das Ziel, weiter auf eine möglichst weitgehende Öffnung der Weltmärkte auf der Basis eines für alle - vom Schwächsten bis zum Stärksten - global gültigen Rechtssystems hinzuarbeiten, wurde festgeschrieben. Dieser weltumspannende Anspruch wurde dadurch unterstrichen, dass die Zahl der Vertragsparteien von unter 100 Ländern Ende der achtziger Jahre inzwischen auf 132 WTO-Mitgliedstaaten gestiegen ist und dass weitere fast	D	LO												
2			8	41	1	700	757	9								
13																
	14	Auch andere Werte														
15	15	Seit kurzem werden freilich kritische Stimmen wieder lauter, die das Heil nicht in einem ungezügelten Liberalismus sehen. Sie werden etwa bestärkt mit dem jüngsten Ausbruch der Finanz- und wirtschaftlichen Strukturkrise in einigen Ländern Südostasiens. Ebenso ist inzwischen die Erkenntnis gewachsen, dass die allein marktwirtschaftlich orientierte Globalisierung durch die weltweite Umverteilung von Besitz und Leistungsbeanspruchung nicht nur Gewinner, sondern auch Verlierer bringt. Mit andern Worten können einzelne Staaten oder sogar ganze Regionen nicht nur einen vergleichweisen Wohlstandsverlust zu anderen Gebieten erleiden, sondern müssen sich eventuell sogar mit einer absoluten Einbusse auseinandersetzen. All dies ist zwar weder wirtschaftstheoretisch noch empirisch neu, doch werden derartige Sachverhalte von den politisch Verantwortlichen nur allzu gerne in den Hintergrund gedrängt.	8	26	1	700	760	9	8	26	1	700	760	9		
3	3		7	12	1522	3	101	7	12	1522	3	101	7	12	1522	3
16	16	Inzwischen machen verschiedene Interessengruppierungen, vor allem aus dem Spektrum der Linksbewegungen, der Umweltaktivisten sowie der eigentlich traditionsbewussten und konservativen Agrarlobby, ihre Bedenken immer stärker geltend. Für sie hat die Weltgemeinschaft noch andere Werte, als möglichst offene Märkte in ihren global wirkenden Ordnungsrahmen einzubeziehen. Stellvertretend für viele sei hier die schweizerische Landwirtschaft erwähnt, die mit europäischer Unterstützung zuhanden der WTO-Ministertagung eine Entschliessung ausgearbeitet hat. Nach Meinung der Agrarlobby muss der Welthandel im Dienste der Gesellschaft stehen und nicht als deren zentraler Organisator wirken. Das multilaterale Regelsystem müsse nicht bloss (kurzfristig) wirtschaftliche Kriterien, sondern fünf weitere Elemente berücksichtigen. Aufgelistet werden ethische, soziale, umweltpolitische, kulturelle und versorgungspolitische Grundsätze, die in die Welthandelsordnung einzubauen sind, wenn sie langfristig Bestand haben will.	8	12	1	700	751	9	8	12	1	700	751	9		
	17	Neue Herausforderung für die WTO														
	18	Für die Landwirtschaft führt die Logik der Globalisierung														

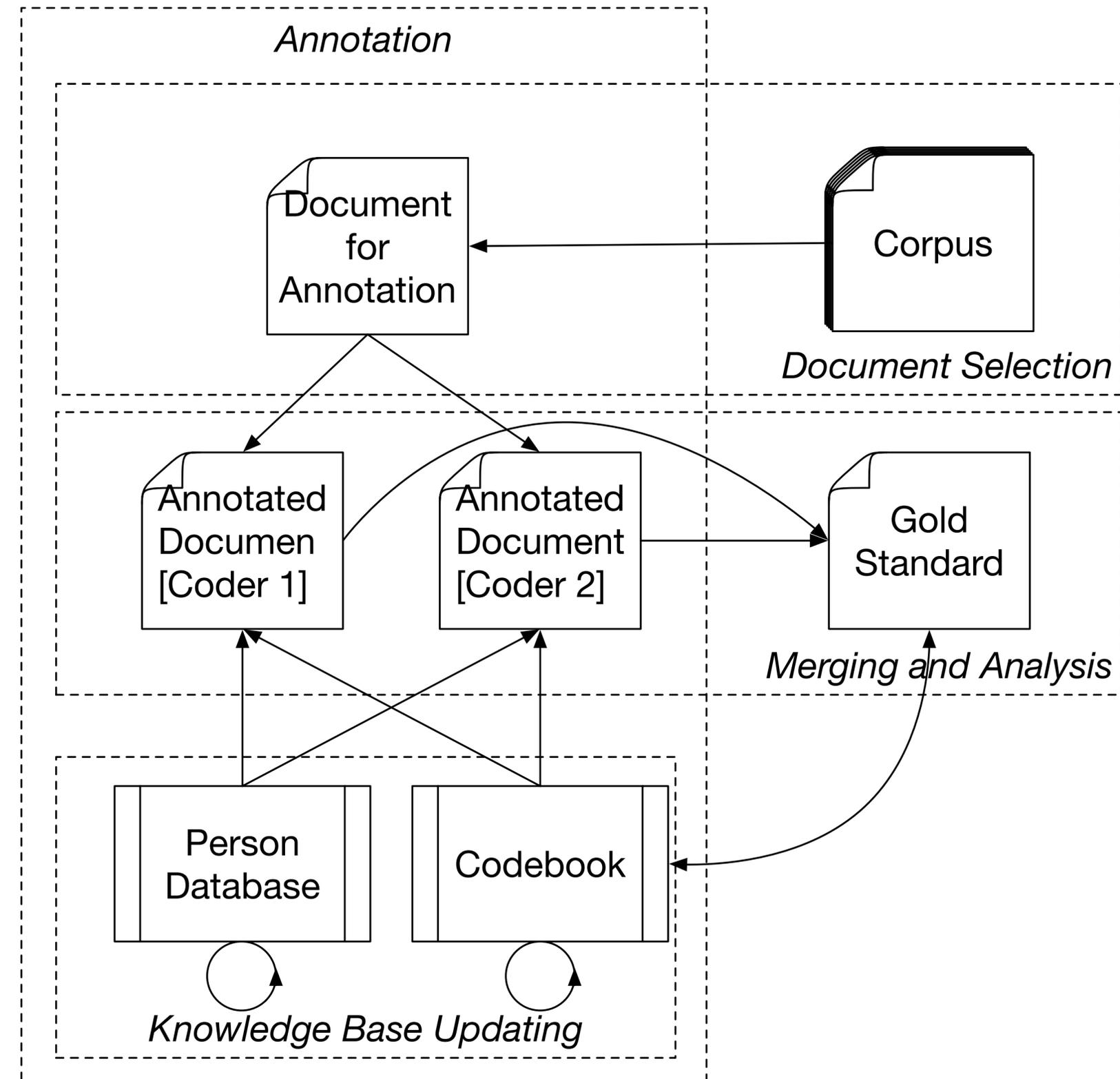
Das Problem

- Workflow
 - ➔ Text-/Artikelauswahl
 - ➔ Claim Annotation
 - ➔ Sprecher-Identifizierung
 - ➔ Claim Klassifizierung
- Daten: Annotation ist **sehr** aufwändig!

Restkodierbuch NZZ		SYST	LO	DR	LP	ML	SP	SP	SYST	LO	DR	LP	ML	SP	SP	
2	1	Neue Zürcher Zeitung	D	26	1	700	757	9								
		Demonstrationen gegen das Welthandelssystem														
	3	Von Wartenweiler, R.														
	4	840 Wörter														
	5	18 Mai 1998														
	6	Neue Zürcher Zeitung														
	7	NEUZZ														
	8	Deutsch														
	9	Besuchen Sie die Website der führenden Schweizer Internationalen Tageszeitung unter http://www.nzz.ch														
	10	Vor der Eröffnung des WTO-Ministertreffens in Genf														
	11	Das multilaterale Welthandelssystem feiert an der diesjährigen WTO-Ministertagung in Genf sein 50-Jahr-Jubiläum. Die Sitzungen im Völkerbundpalast sind entsprechend den Zufälligkeiten der Präsenz der verschiedenen Persönlichkeiten und zulasten des festlichen Charakters kunterbunt zusammengestellt. Das frühzeitig aufgebaute Sicherheitssystem ist bereits im Vorfeld der Zusammenkunft von Demonstrationen gegen die Globalisierung und gegen den liberalen Welthandel getestet worden.	D	26	1	700	757	9								
4	12	Ww. Genf, 17. Mai														
	13	Im Zuge des Zusammenbruchs des Systems der kommunistischen Planwirtschaft und des Falls der Berliner Mauer sind die neunziger Jahre mit einer Euphorie über die Geltungskraft marktwirtschaftlicher Prinzipien «eingeläutet» worden. Von diesem Elan profitierte auch die Uruguay-Runde, mit deren Abschluss letztlich das 1948 geschaffene Provisorium des Allgemeinen Zoll- und Handelsabkommens (Gatt) am 1. Januar 1995 durch die auf Dauer angelegte Institution der Welthandelsorganisation (WTO) abgelöst wurde. Das Ziel, weiter auf eine möglichst weitgehende Öffnung der Weltmärkte auf der Basis eines für alle - vom Schwächsten bis zum Stärksten - global gültigen Rechtssystems hinzuarbeiten, wurde festgeschrieben. Dieser weltumspannende Anspruch wurde dadurch unterstrichen, dass die Zahl der Vertragsparteien von unter 100 Ländern Ende der achtziger Jahre inzwischen auf 132 WTO-Mitgliedstaaten gestiegen ist und dass weitere fast	D	26	1	700	757	9								
2		Sinnhaftigkeit														
13																
	14	Auch andere Werte														
15	15	Seit kürzlich werden freilich kritische Stimmen wieder lauter, die das Heil nicht in einem ungezügelten Liberalismus sehen. Sie werden etwa bestärkt mit dem jüngsten Ausbruch der Finanz- und wirtschaftlichen Strukturkrise in einigen Ländern Südostasiens. Ebenso ist inzwischen die Erkenntnis gewachsen, dass die allein marktwirtschaftlich orientierte Globalisierung durch die weltweite Umverteilung von Besitz und Leistungsbeanspruchung nicht nur Gewinner, sondern auch Verlierer bringt. Mit andern Worten können einzelne Staaten oder sogar ganze Regionen nicht nur einen vergleichweisen Wohlstandsverlust zu anderen Gebieten erleiden, sondern müssen sich eventuell sogar mit einer absoluten Einbusse auseinandersetzen. All dies ist zwar weder wirtschaftstheoretisch noch empirisch neu, doch werden derartige Sachverhalte von den politisch Verantwortlichen nur allzu gerne in den Hintergrund gedrängt.														
3	16	Inzwischen machen verschiedene Interessengruppierungen, vor allem aus dem Spektrum der Linksbewegungen, der Umweltaktivisten sowie der eigentlich traditionsbewussten und konservativen Agrarlobby, ihre Bedenken immer stärker geltend. Für sie hat die Weltgemeinschaft noch andere Werte, als möglichst offene Märkte in ihren global wirkenden Ordnungsrahmen einzubeziehen. Stellvertretend für viele sei hier die schweizerische Landwirtschaft erwähnt, die mit europäischer Unterstützung zuhanden der WTO-Ministertagung eine Entschliessung ausgearbeitet hat. Nach Meinung der Agrarlobby muss der Welthandel im Dienste der Gesellschaft stehen und nicht als deren zentraler Organisator wirken. Das multilaterale Regelsystem müsse nicht bloss (kurzfristig) wirtschaftliche Kriterien, sondern fünf weitere Elemente berücksichtigen. Aufgelistet werden ethische, soziale, umweltpolitische, kulturelle und versorgungspolitische Grundsätze, die in die Welthandelsordnung einzubauen sind, wenn sie langfristig Bestand haben will.														
	17	Neue Herausforderung für die WTO														
	18	Für die Landwirtschaft führt die Logik der Globalisierung														

Die Lösung?

- (Semi-)automatische Annotation mit Hilfe von NLP und maschinellem Lernen
- **Workflow**
 - ➔ Corpuserstellung [Import, Tokenisierung]
 - ➔ Dokumentenauswahl [key words, topic models]
 - ➔ Annotation [manuell + KI-unterstützt]
 - Claims [Codebuch]
 - Akteure [Wissensdatenbank; z.B. Wikidata]
 - KI: Claim-Erkennung und Klassifizierung: BERT
 - ➔ Zusammenführung und Evaluation [Goldstandard]
 - ➔ Wissensdatenbank-Update
 - ➔ Datenexport



Annotations Tool

Annotator:innen Ansicht

semi-automatische Annotation (Model 6)

abschliessen

<p>-> 100 [AI] : CDU / CSU Horst Seehofer droht der Kanzlerin mit Konsequenzen , sollte die Zahl der Flüchtlinge nicht sinken .</p>	<p>show 100 102 [user] : CDU / CSU Horst Seehofer droht der Kanzlerin mit Konsequenzen , sollte die Zahl der Flüchtlinge nicht sinken .</p> <p>1 20151028 edit</p>
	<p>show 800 801 [user] : Vor zwei Wochen drohte Bayerns Ministerpräsident bereits mit einer Verfassungsklage, sollte die – von seiner CSU übrigens mit geführte – Bundesregierung nicht für die Einhaltung geltender Gesetze sorgen.</p> <p>1 20151028 edit</p>
<p>-> 100 [AI] : Nun hat Seehofer der Bundesregierung am Dienstag ein Ultimatum gestellt : Bis zum Sonntag soll Angela Merkel die Zuwanderung von Flüchtlingen begrenzen .</p>	<p>show 100 102 [user] : Nun hat Seehofer der Bundesregierung am Dienstag ein Ultimatum gestellt : Bis zum Sonntag soll Angela Merkel die Zuwanderung von Flüchtlingen begrenzen .</p> <p>1 20151028 edit</p>
<p>show -> 500 800 [AI] : In der Regierung sei " ständige gegenseitige Information notwendig " .</p>	
<p>show -> 800 [AI] : " Wir sind da ganz fest und hart in der Sache , aber ich möchte jetzt erst die Gespräche führen . "</p>	

KI Vorschläge

Metadaten

MetaDaten [id: 143569] [Datum: 29.10.2015] [Ressort: Schwerpunkt] [Quelle: taz] [Seite: 2] [Art:TAZ-Bericht] taz2015/2015_10_29/art00762781.xml]

Ultimatum mit was denn bloß

CDU/CSU Horst Seehofer droht der Kanzlerin mit Konsequenzen, sollte die Zahl der Flüchtlinge nicht sinken. Doch das Krawallpotenzial des CSU-Vorsitzenden ist begrenzt

Anja Maier

Named entity detection

Beim Thema Flüchtlinge ist Horst Seehofer (CSU) mittlerweile ein Meister der Ankündigungen. Vor zwei Wochen drohte Bayerns Ministerpräsident bereits mit einer Verfassungsklage, sollte die – von seiner CSU übrigens mit geführte – Bundesregierung nicht für die Einhaltung geltender Gesetze sorgen. Die Klage blieb bislang bekanntlich aus. An der großen Anzahl von Flüchtlingen, die tagtäglich von Österreich aus in Bayern eintreffen, hat sich seither aber nichts geändert – eher sind es noch mehr geworden.

Nun hat Seehofer der Bundesregierung am Dienstag ein Ultimatum gestellt: Bis zum Sonntag soll Angela Merkel die Zuwanderung von Flüchtlingen begrenzen.

Und was, wenn (wie jedem klar sein müsste) das nicht geht? Will Horst Seehofer dann die erschöpften Flüchtlingsfamilien an der Grenze zurück Richtung Syrien schicken? Will er einen Zaun zu Österreich errichten lassen? Wird er die Erstaufnahmeeinrichtungen in Bayern bis auf Weiteres schließen? Wohl kaum.

Annotations Tool

Annotator:innen Ansicht

semi-automatische Annotation (Model 6)

abschliessen

-> 100 [AI] : CDU / CSU Horst Seehofer droht der Kanzlerin mit Konsequenzen , sollte die Zahl der Flüchtlinge nicht sinken .

show 100 102 [user] : CDU / CSU Horst Seehofer droht der Kanzlerin mit Konsequenzen , sollte die Zahl der Flüchtlinge nicht sinken .

1 20151028 edit

show 800 801 [user] : Vor zwei Wochen drohte Bayerns Ministerpräsident bereits mit einer Verfassungsklage, sollte die – von seiner CSU übrigens mit geführte – Bundesregierung nicht für die Einhaltung geltender Gesetze sorgen.

1 20151028 edit

-> 100 [AI] : Nun hat Seehofer der Bundesregierung am Dienstag ein Ultimatum gestellt : Bis zum Sonntag soll Angela Merkel die Zuwanderung von Flüchtlingen begrenzen .

show 100 102 [user] : Nun hat Seehofer der Bundesregierung am Dienstag ein Ultimatum gestellt : Bis zum Sonntag soll Angela Merkel die Zuwanderung von Flüchtlingen begrenzen .

1 20151028 edit

show -> 500 800 [AI] : In der Regierung sei " ständige gegenseitige Information notwendig " .

show -> 800 [AI] : " Wir sind da ganz fest und hart in der Sache , aber ich möchte jetzt erst die Gespräche führen . "

Pop-up Annotationsfenster

MetaDaten [id: 143569] [Datum: 29.10.2015] [Ressort: Schwerpunkt] [Quelle: taz] [Seite: 2] [Art:TAZ-Bericht] taz2015/2015_10_29/art00762781.xml]

Ultimatum mit was denn bloß

CDU/CSU Horst Seehofer droht der Kanzlerin mit Konsequenzen, sollte die Zahl der Flüchtlinge nicht sinken.

Claim Frame Actor new PER/ORG

CPOS: positiv CDATE: 28.10.2015
 grobe Zeitangabe

ClaimSummary

- Steuerung von Migration (100)
 - kontrollierte Zuwanderung (101)
 - Begrenzung/Obergrenze (102)
 - Abschottung/Zuwanderungsstopp (104)
 - Grenzkontrollen (105)
 - Grenzverteidigung (106)
 - Zaun (107)
 - Einwanderungsgesetz (108)
 - Kampf gegen Schlepper (109)
 - Asylrecht (110)
 - Seenotrettung (111)
 - Aktuelle Migrationspolitik (190)
 - Allgemein (199)

Cancel Save

Comments...

Und was, wenn (wie jedem klar sein müsste) das nicht geht? Will Horst Seehofer dann die erschöpften Flüchtlingsfamilien an der Grenze zurück Richtung Syrien schicken? Will er einen Zaun zu Österreich errichten lassen? Wird er die Erstaufnahmeeinrichtungen in Bayern bis auf Weiteres schließen? Wohl kaum.

Annotations Tool Goldstandard

show -> 800801 Vor zwei Wochen drohte Bayerns Ministerpräsident bereits mit einer Verfassungsklage, sollte die – von seiner CSU übrigens mit geführte – Bundesregierung nicht für die Einhaltung geltender Gesetze sorgen.
1 20151028

show -> 800809 Vor zwei Wochen drohte Bayerns Ministerpräsident bereits mit einer Verfassungsklage, sollte die – von seiner CSU übrigens mit geführte – Bundesregierung nicht für die Einhaltung geltender Gesetze sorgen.
1 20151028

show -> 800801 Vor zwei Wochen drohte Bayerns Ministerpräsident bereits mit einer Verfassungsklage, sollte die – von seiner CSU übrigens mit geführte – Bundesregierung nicht für die Einhaltung geltender Gesetze sorgen.
1 20151028

show -> 800801 Vor zwei Wochen drohte Bayerns Ministerpräsident bereits mit einer Verfassungsklage, sollte die – von seiner CSU übrigens mit geführte – Bundesregierung nicht für die Einhaltung geltender Gesetze sorgen.
1 20151028

show -> 100102 Nun hat Seehofer der Bundesregierung am Dienstag ein Ultimatum gestellt: Bis zum Sonntag soll Angela Merkel die Zuwanderung von Flüchtlingen begrenzen.
1 20151028

show -> 100102 Nun hat Seehofer der Bundesregierung am Dienstag ein Ultimatum gestellt: Bis zum Sonntag soll Angela Merkel die Zuwanderung von Flüchtlingen begrenzen.
1 20151028

show -> 100 Nun hat Seehofer der Bundesregierung am Dienstag ein Ultimatum gestellt : Bis zum Sonntag soll Angela Merkel die Zuwanderung von Flüchtlingen begrenzen .
1 20151028

800801 [0] : Vor zwei Wochen drohte Bayerns Ministerpräsident bereits mit einer Verfassungsklage, sollte die – von seiner CSU übrigens mit geführte – Bundesregierung nicht für die Einhaltung geltender Gesetze sorgen.
WVL
per_Horst Seehofer
1 20151014 loeschen edit

Annotations von
verschiedenen
Kodierern

KI Vorschlag

100102 [0] : Nun hat Seehofer der Bundesregierung am Dienstag ein Ultimatum gestellt: Bis zum Sonntag soll Angela Merkel die Zuwanderung von Flüchtlingen begrenzen.
WVL

Ausgewählte Goldstandard Annotation

MetaDaten [id: 143569] [Datum: 29.10.2015] [Ressort: Schwerpunkt] [Quelle: taz] [Seite: 2] [Art:TAZ-Bericht] taz2015/2015_10_29/art00762781.xml]

Ultimatum mit was denn bloß

CDU/CSU Horst Seehofer droht der Kanzlerin mit Konsequenzen, sollte die Zahl der Flüchtlinge nicht sinken. Doch das Krawallpotenzial des CSU-Vorsitzenden ist begrenzt

Aus Berlin

Anja Maier

Beim Thema Flüchtlinge ist Horst Seehofer (CSU) mittlerweile ein Meister der Ankündigungen. Vor zwei Wochen drohte Bayerns Ministerpräsident bereits mit einer Verfassungsklage, sollte die – von seiner CSU übrigens mit geführte – Bundesregierung nicht für die Einhaltung geltender Gesetze sorgen. Die Klage blieb bislang bekanntlich aus. An der großen Anzahl von Flüchtlingen, die tagtäglich von Österreich aus in Bayern eintreffen, hat sich seither aber nichts geändert – eher sind es noch mehr geworden.

Nun hat Seehofer der Bundesregierung am Dienstag ein Ultimatum gestellt: Bis zum Sonntag soll Angela Merkel die Zuwanderung von Flüchtlingen begrenzen.

Und was, wenn (wie jedem klar sein müsste) das nicht geht? Will Horst Seehofer dann die erschöpften Flüchtlingsfamilien an der Grenze zurück Richtung Syrien schicken? Will er einen Zaun zu Österreich errichten lassen? Wird er die Erstaufnahmeeinrichtungen in Bayern bis auf Weiteres schließen? Wohl kaum.

Ebenso unwahrscheinlich ist, dass Seehofer die Fraktionsgemeinschaft seiner CSU mit der CDU in Berlin aufkündigt oder von dort gar seine drei Bundesminister abzieht.

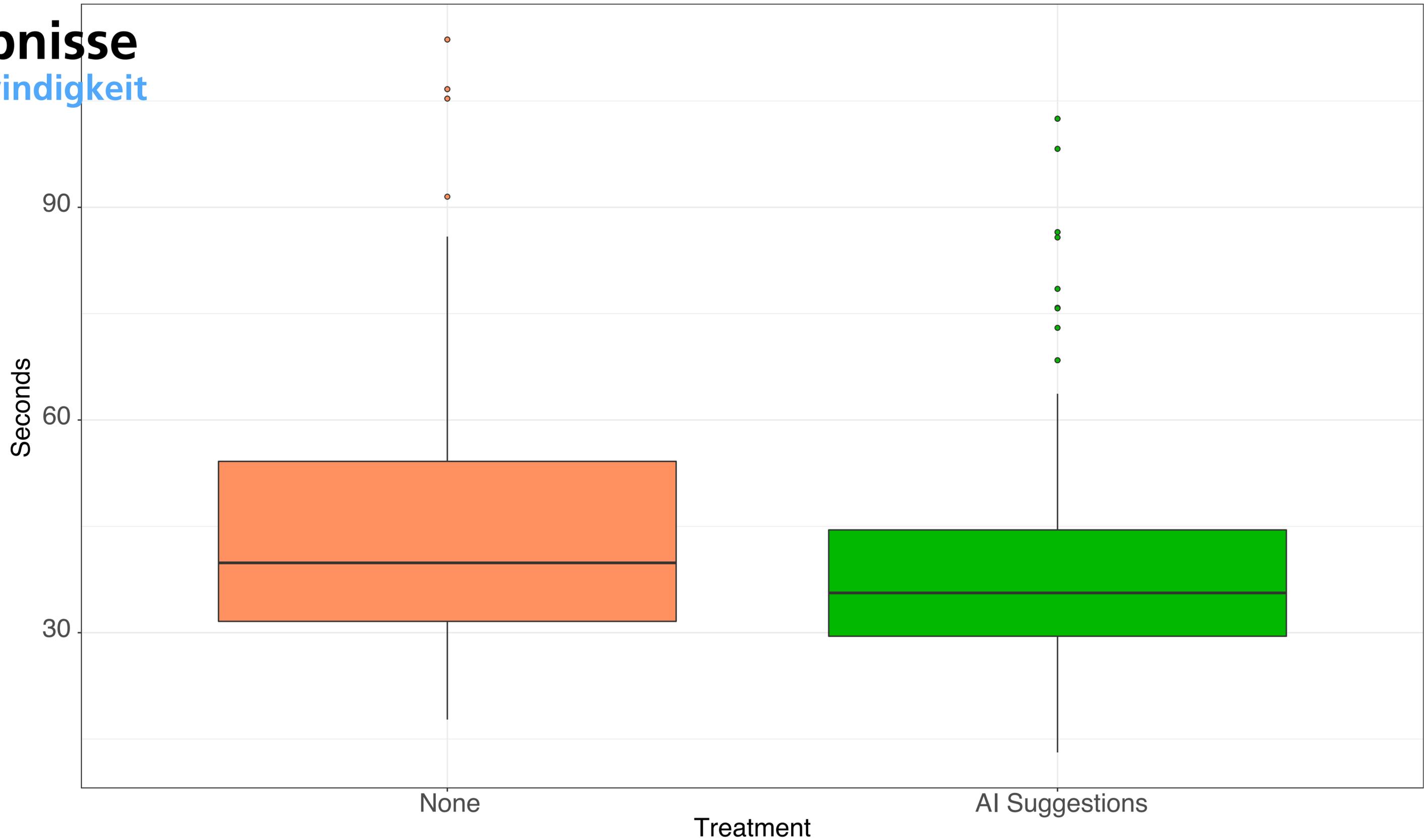
Experiment

- 6 Erfahrene Annotierer:innen
- 2 Gruppen
 - ➔ Treatment = Vorschläge von der KI
 - ➔ Kein Treatment = manuelle Annotation ohne KI-Vorschläge
- 4 Runden à 10 Artikel
- Messung
 - ➔ durchschnittliche Zeit für die Annotation eines Artikels
 - ➔ Precision, Recall

	Gruppe A	Gruppe B
Runde 1	Treatment	Kein Treatment
Runde 2	Kein Treatment	Treatment
Runde 3	Kein Treatment	Treatment
Runde 4	Treatment	Kein Treatment

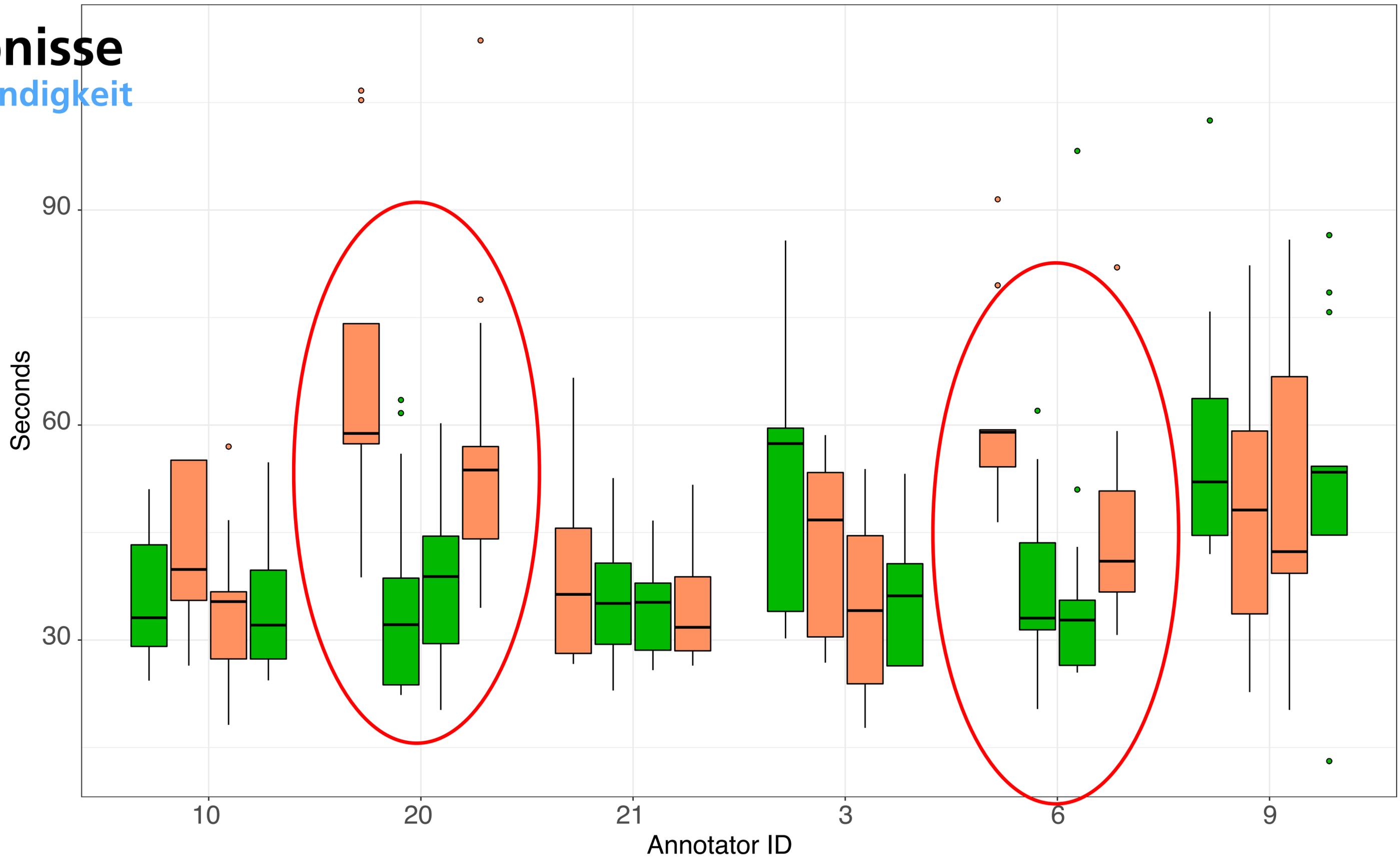
Ergebnisse

Geschwindigkeit



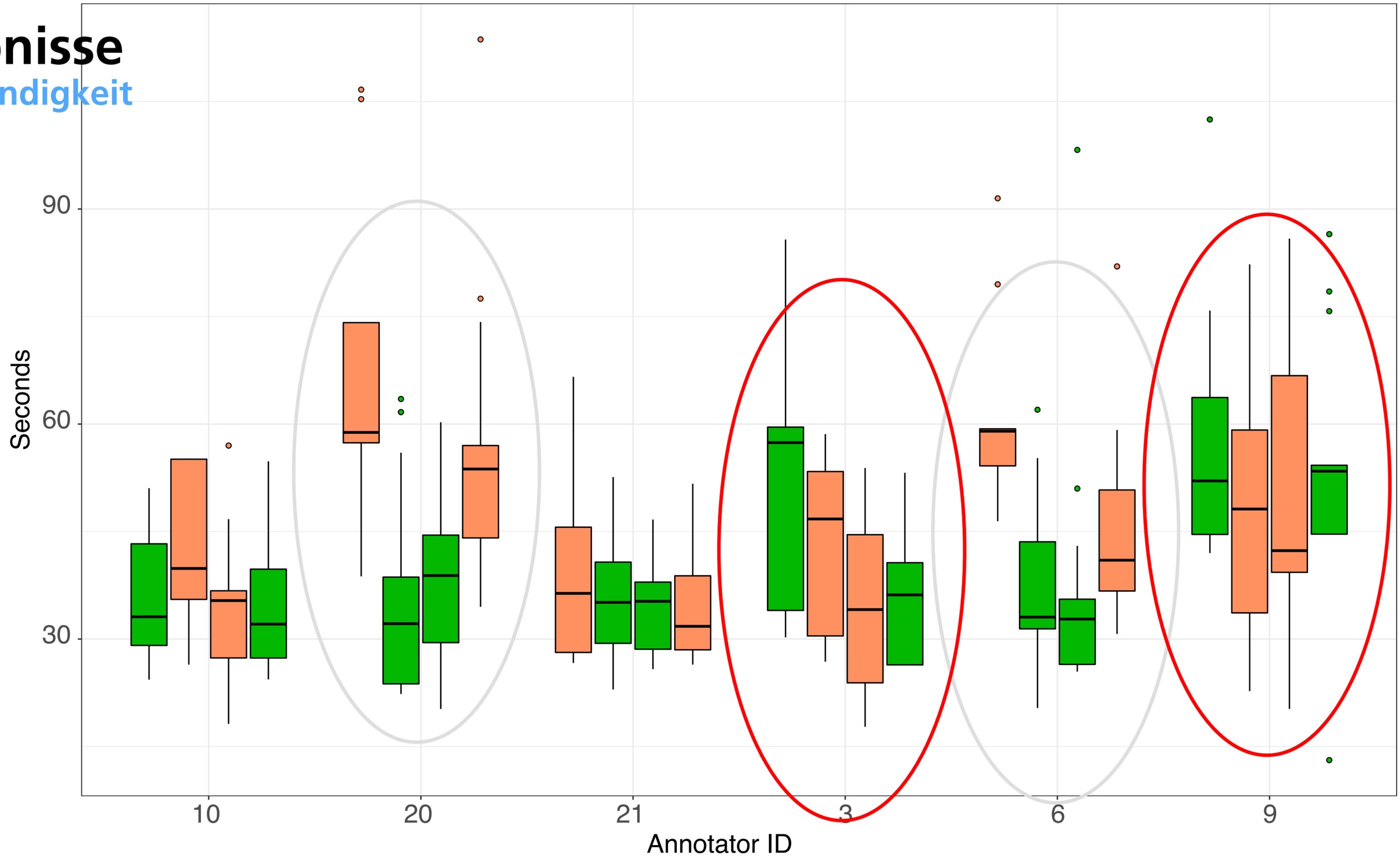
Ergebnisse

Geschwindigkeit



Ergebnisse

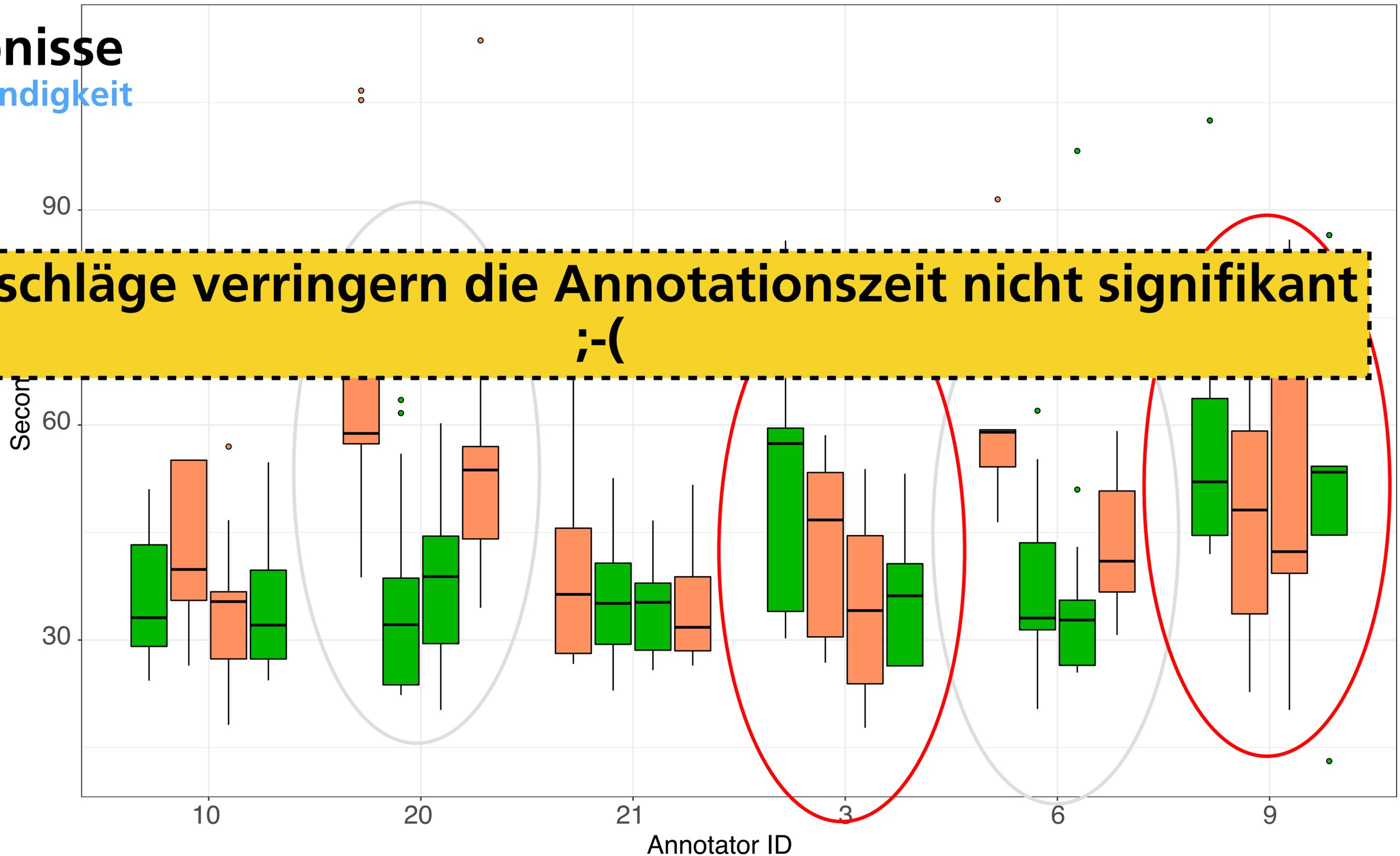
Geschwindigkeit



Ergebnisse

Geschwindigkeit

KI-Vorschläge verringern die Annotationszeit nicht signifikant :-)



Ergebnisse

Precision & Recall

Treatment	Recall	Precision	F1
ohne	74.11	81.85	76.73
mit KI Vorschlägen	79.99	81.41	80.10

Ergebnisse

Precision & Recall

Treatment	Recall	Precision	F1
ohne	74.11	81.85	76.73
mit KI Vorschlägen	79.99	81.41	80.10

Besserer Recall und insgesamt bessere Ergebnisse

Ergebnisse

Precision & Recall

Treatment	Recall	Precision	F1
ohne	74.11	81.85	76.73
mit KI Vorschlägen	79.99	81.41	80.10

Besserer Recall und insgesamt bessere Ergebnisse

KI Performance

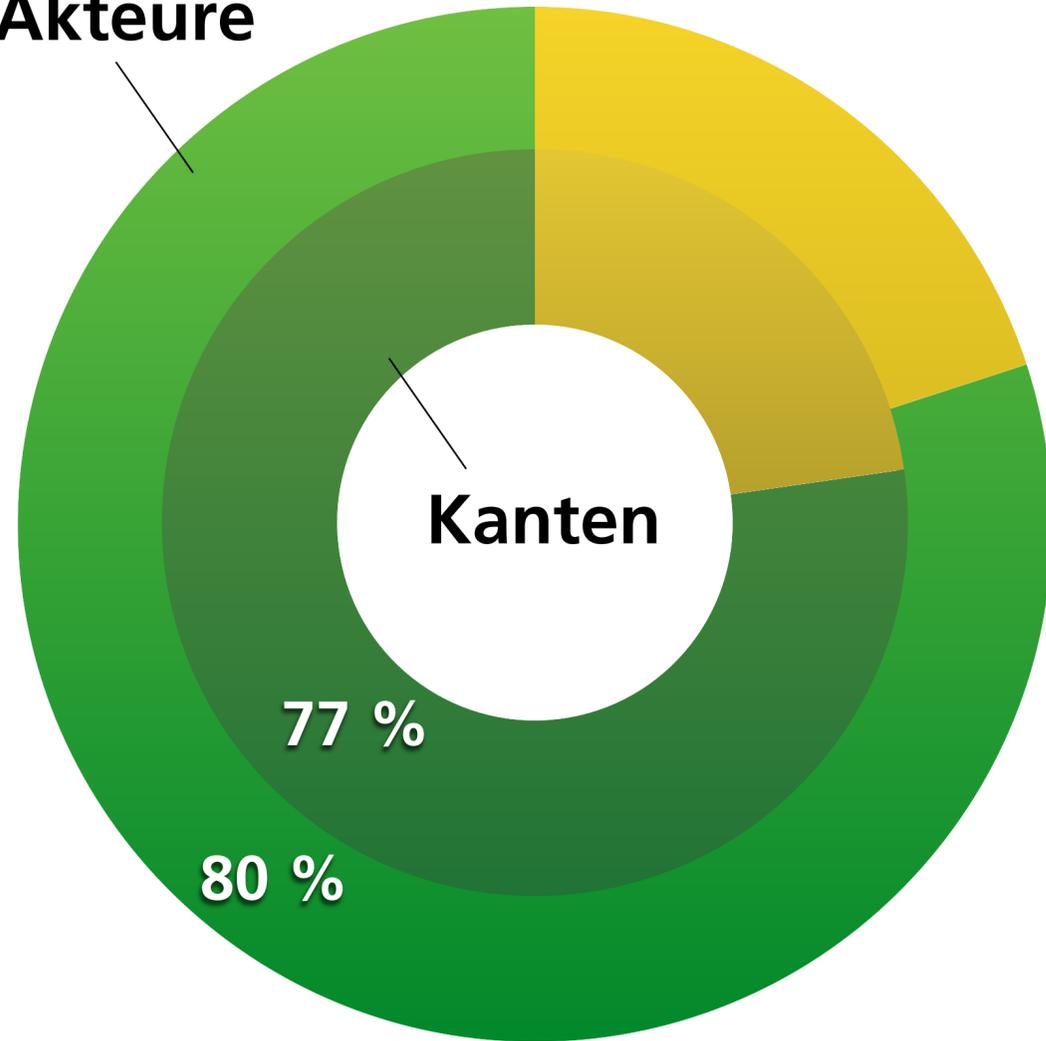
70.9

73.1

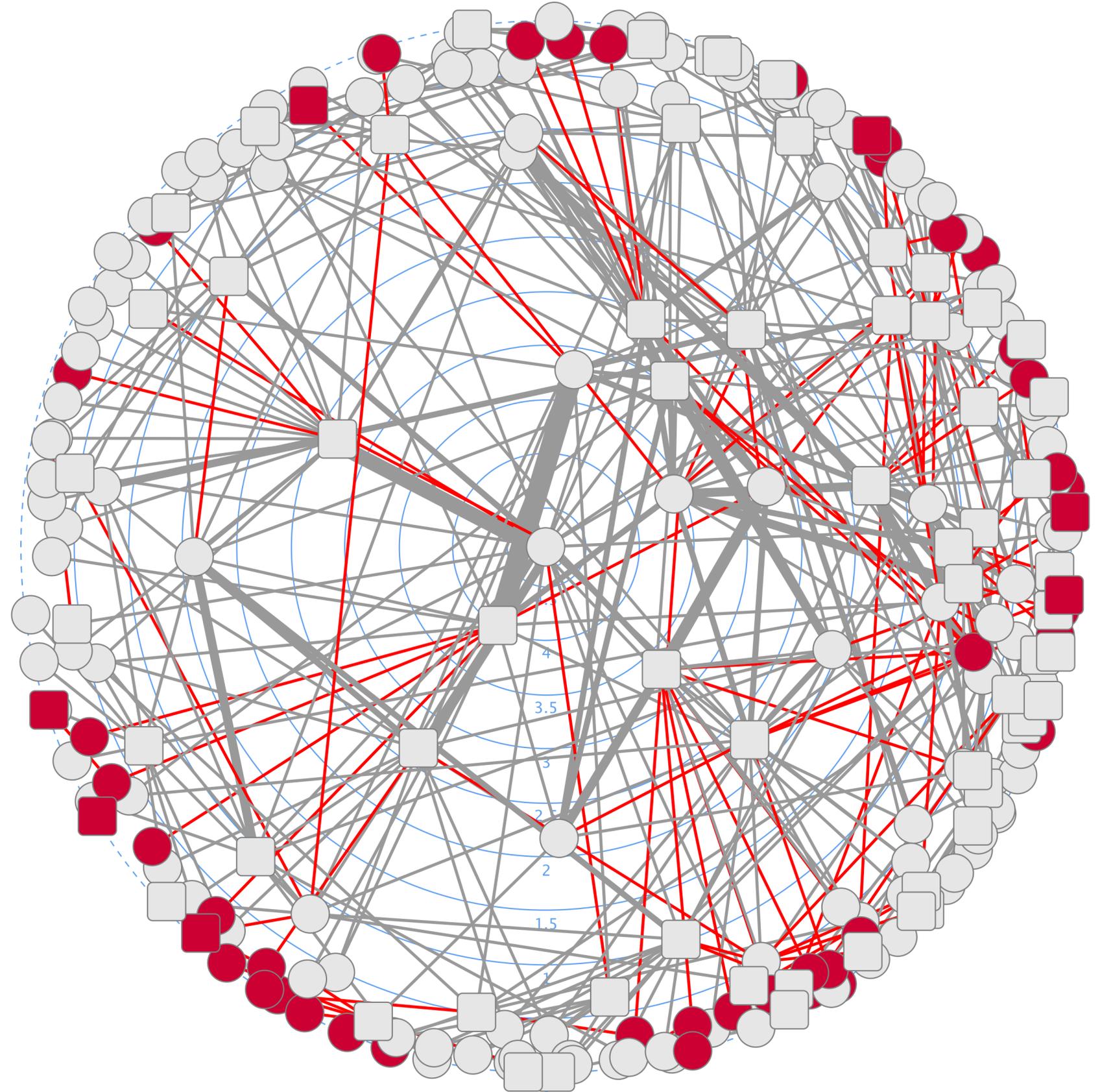
71.9

KI Vorschläge

Akteure

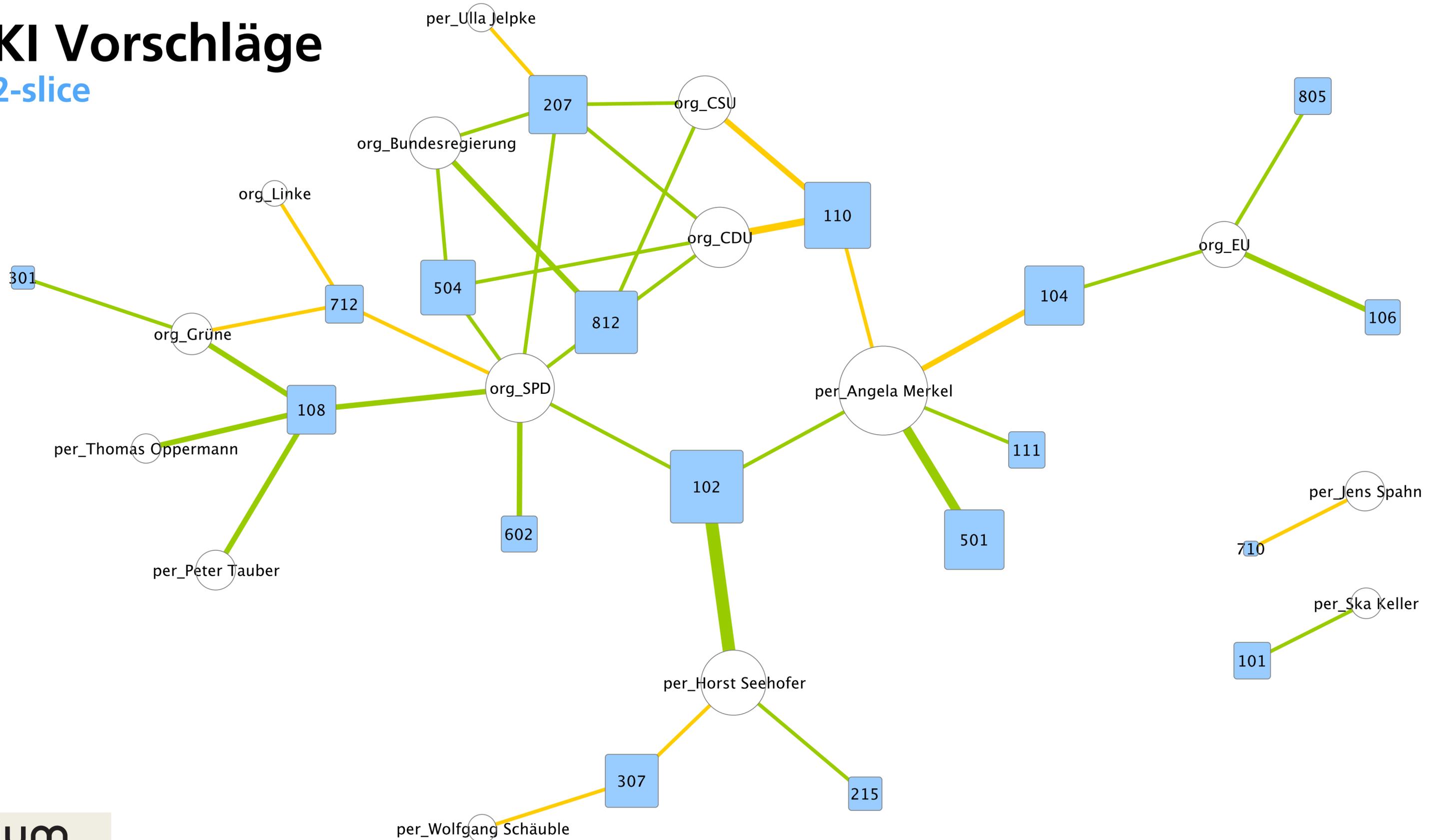


Precision = 70.9
Recall = 73.1
F1 = 71.9



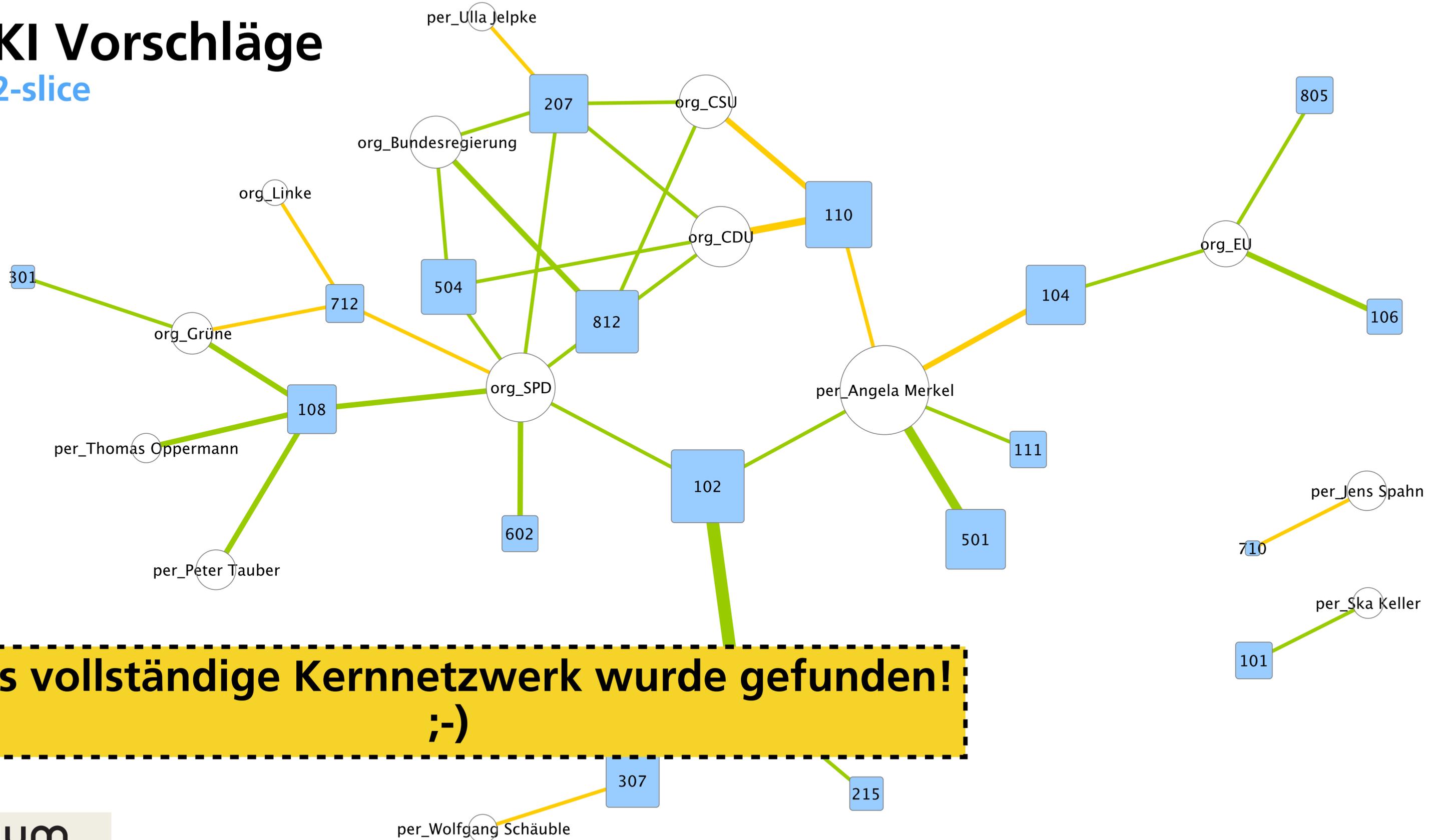
KI Vorschläge

2-slice



KI Vorschläge

2-slice



Schlussfolgerungen

- ML verbessert die Qualität der Annotation (recall & F1), aber der Geschwindigkeitsgewinn ist nur marginal
- Ein KI Pseudo-Coder ist nicht sehr zuverlässig

Schlussfolgerungen

- ML verbessert die Qualität der Annotation (recall & F1), aber der Geschwindigkeitsgewinn ist nur marginal
- Ein KI Pseudo-Coder ist nicht sehr zuverlässig

ABER

- Ein KI Pseudo-Coder erkennt den Kern des Diskursnetzwerks sehr zuverlässig
- Damit lassen sich Informationen über zentrale Akteure und ihre Forderungen (fast) vollautomatisiert extrahieren.

Methoden-Mix / Methodenintegration

- ML ist auf kuratierte, manuell annotierte Datensätze angewiesen
- KI-basierte Methoden der quantitativen Textanalyse können qualitative Methoden nicht vollständig ersetzen
- Sie haben aber das Potenzial den Annotationsaufwand für große Datensätze drastisch zu reduzieren

➔ Weitergehende Informationen auf <https://www.mardy-project.eu>

Article

Integrating Manual and Automatic Annotation for the Creation of Discourse Network Data Sets

Sebastian Haunss^{1,*}, Jonas Kuhn², Sebastian Padó², Andre Blessing², Nico Blokker¹, Erenay Dayanik² and Gabriella Lapesa²

¹ Research Center on Inequality and Social Policy, University of Bremen, 28359 Bremen, Germany; E-Mails: sebastian.haunss@uni-bremen.de (S.H.), blokker@uni-bremen.de (N.B.)

² Institute for Natural Language Processing, University of Stuttgart, 70569 Stuttgart, Germany; E-Mails: jonas.kuhn@ims.uni-stuttgart.de (J.K.), sebastian.pado@ims.uni-stuttgart.de (S.P.), andre.blessing@ims.uni-stuttgart.de (A.B.), erenay.dayanik@ims.uni-stuttgart.de (E.D.), gabriella.lapesa@ims.uni-stuttgart.de (G.L.)

* Corresponding author

Submitted: 30 October 2019 | Accepted: 23 March 2020 | Published: 2 June 2020

Abstract

This article investigates the integration of machine learning in the political claim annotation workflow with the goal to partially automate the annotation and analysis of large text corpora. It introduces the MARDY annotation environment and presents results from an experiment in which the annotation quality of annotators with and without machine learning based annotation support is compared. The design and setting aim to measure and evaluate: a) annotation speed; b) annotation quality; and c) applicability to the use case of discourse network generation. While the results indicate only slight increases in terms of annotation speed, the authors find a moderate boost in annotation quality. Additionally, with the help of manual annotation of the actors and filtering out of the false positives, the machine learning based annotation suggestions allow the authors to fully recover the core network of the discourse as extracted from the articles annotated during the experiment. This is due to the redundancy which is naturally present in the annotated texts. Thus, assuming a research focus not on the complete network but the network core, an AI-based annotation can provide reliable information about discourse networks with much less human intervention than compared to the traditional manual approach.

Keywords

annotation; automation; discourse networks; machine learning; migration discourse

Issue

This article is part of the issue “Policy Debates and Discourse Network Analysis” edited by Phillip Leifeld (University of Essex, UK).

© 2020 by the authors; licensee Cogitatio (Lisbon, Portugal). This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

1. Introduction

Discourse network analysis (DNA) offers a conceptual framework for the analysis of discourse structures and dynamics. Numerous DNA studies have shown that the network perspective on political discourse offers insights that go beyond traditional policy analyses and qualitative discourse studies (Haunss, 2017; Leifeld, 2016; Nagel & Satoh, 2019; Wang & Wang, 2017). In principle, modelling the development of political debates as dynamic

discourse networks may enable us to identify recurring mechanisms that drive the development of political debates and to distinguish between network effects and actor attribute effects. Unfortunately, the creation of dynamic discourse network data sets is extremely time- and labour-intensive and therefore poses a serious barrier for this kind of analysis.

In this article, we present the first results from a research project in which we investigate annotation workflows that integrate machine learning to partially auto-

Improving Neural Political Statement Classification with Class Hierarchical Information

Erenay Dayanik¹, André Blessing¹, Nico Blokker², Sebastian Haunss², Jonas Kuhn¹, Gabriella Lapesa¹, and Sebastian Padó¹

¹IMS, University of Stuttgart, Germany

²SOCIUM, University of Bremen, Germany

Abstract

Many tasks in text-based computational social science (CSS) involve the classification of political statements into categories based on a domain-specific codebook. In order to be useful for CSS analysis, these categories must be fine-grained. The typically skewed distribution of fine-grained categories, however, results in a challenging classification problem on the NLP side. This paper proposes to make use of the *hierarchical relations among categories* typically present in such codebooks: e.g., *markets* and *taxation* are both subcategories of *economy*, while *borders* is a subcategory of *security*. We use these ontological relations as prior knowledge to establish additional constraints on the learned model, thus improving performance overall and in particular for infrequent categories. We evaluate several lightweight variants of this intuition by extending state-of-the-art transformer-based text classifiers on two datasets and multiple languages. We find the most consistent improvement for an approach based on regularization.

1 Introduction

The argumentative or discursive turn in policy analysis and political science more generally has long established the value of textual sources for the analysis of politics and policies (Fischer and Forester, 1993). Traditionally, data sources such as interviews or newspaper reports were annotated using various methods of qualitative text analysis (Wagenaar, 2011; Mayring, 2019). At the heart of this analysis is always a *codebook*, i.e., guidelines that map actual statements or textual passages to the abstract concepts relevant for the respective research.

Categories in codebooks are almost always arranged *hierarchically*, with *fine-grained* categories being grouped together into supercategories that are often, but not always, more abstract. Fine-grained categories are generally generated inductively from the analyzed texts in an iterative pro-

cess of summarizing and abstracting from the original text, while the supercategories are deductively generated from existing knowledge of the relevant policy field and from theoretical and conceptual findings of prior research. For example, the codebook of the long-running Comparative Manifesto Project (CMP), which analyzes party manifestos across several countries, includes 7 supercategories (such as *external relations* or *economy*) with 56 subcategories: for *economy*, among others, *free market*, *market regulation*, *economic goals*, etc. (Merz et al., 2016; Werner et al., 2011). Here, supercategories represent the separation of policy fields that is reflected in political institutions, e.g., ministries. Fine-grained, hierarchical schemes help researchers both with data annotation and with analysis. Annotation is often easier when the annotation decision is (implicitly) first based on a supercategory and then on fine-grained subcategories. For analysis, supercategories structure the annotated material according to different levels of abstraction, thereby supporting interpretation and modeling.

While such a hierarchical process a natural choice in manual annotation, the situation is different when we move to (semi)-automatic analysis in NLP: due to the large number of fine-grained subcategories, the available data is distributed among many categories. In addition, most categories are infrequently attested, since categories typically show a skewed distribution. This makes for a difficult classification problem, and existing prediction studies have often only addressed the more coarse-grained supercategory level (Glavaš et al., 2017a; Subramanian et al., 2018; Padó et al., 2019).

In this study, we ask whether we can use the hierarchical structure of political science codebooks to our advantage: knowing that two subcategories (as *free market* and *market regulation*) belong to the same supercategory (*economy*) could lead us to expect that the representations learned for these categories should be more similar to one another