

Möglichkeiten und Grenzen des Weblyzards zur automatisierten Erfassung der Sicherheitsheitswahrnehmungen in Österreich

Eine Fallstudie und ein Plädoyer für eine kritische Soziologie im Spannungsfeld zwischen ökonomischer Vermarktung und politischer Anwendung von Big Data

Martin Weichbold und Alexander Seymer
gemeinsam mit Wolfgang Aschauer (PI), Thomas Herdin,
Patric Messner, Lena Stöllinger

CAU Kiel 2018
Masse statt Klasse?

Entstehungsprozesse, Qualitätsprobleme und Verwendungsmöglichkeiten
von prozessgenerierten Daten im digitalen Zeitalter

15. Februar 2018

Aufbau des Vortrags

1. Kurze Vorstellung des Projekts und der Akteure

Aufbau des Vortrags

1. Kurze Vorstellung des Projekts und der Akteure
2. Unsere Projektstätigkeit im Rahmen des „Forestrat-Cockpits“

Aufbau des Vortrags

1. Kurze Vorstellung des Projekts und der Akteure
2. Unsere Projektstätigkeit im Rahmen des „Forestrat-Cockpits“
3. Vorstellung des Weblyzards

Aufbau des Vortrags

1. Kurze Vorstellung des Projekts und der Akteure
2. Unsere Projektstätigkeit im Rahmen des „Forestrat-Cockpits“
3. Vorstellung des Weblyzards
4. Vorstellung der Evaluationsstudie („Mensch vs. Maschine“)

Aufbau des Vortrags

1. Kurze Vorstellung des Projekts und der Akteure
2. Unsere Projektstätigkeit im Rahmen des „Forestrat-Cockpits“
3. Vorstellung des Weblyzards
4. Vorstellung der Evaluationsstudie („Mensch vs. Maschine“)
5. Qualitätskriterien zum Umgang mit Medienquellen

Aufbau des Vortrags

1. Kurze Vorstellung des Projekts und der Akteure
2. Unsere Projektstätigkeit im Rahmen des „Forestrat-Cockpits“
3. Vorstellung des Weblyzards
4. Vorstellung der Evaluationsstudie („Mensch vs. Maschine“)
5. Qualitätskriterien zum Umgang mit Medienquellen
6. Einordnung der Evaluation in den Workshop

Projekt und beteiligte Akteure

- FFG-finanziertes Projekt im Rahmen der KIRAS-Programmlinie

Projekt und beteiligte Akteure

- FFG-finanziertes Projekt im Rahmen der KIRAS-Programmlinie
- Das Konsortium bestand aus:
 - Bedarfsträgern (BMLVS und BMI)

Projekt und beteiligte Akteure

- FFG-finanziertes Projekt im Rahmen der KIRAS-Programmlinie
- Das Konsortium bestand aus:
 - Bedarfsträgern (BMLVS und BMI)
 - Wirtschaftsunternehmen (Repuco und RISE)

Projekt und beteiligte Akteure

- FFG-finanziertes Projekt im Rahmen der KIRAS-Programmlinie
- Das Konsortium bestand aus:
 - Bedarfsträgern (BMLVS und BMI)
 - Wirtschaftsunternehmen (Repuco und RISE)
 - Wissenschaft (Universität Salzburg und Modul Universität Wien)

Projekt und beteiligte Akteure

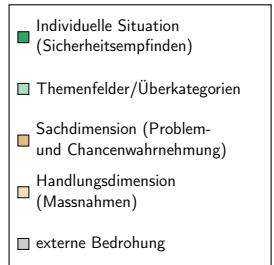
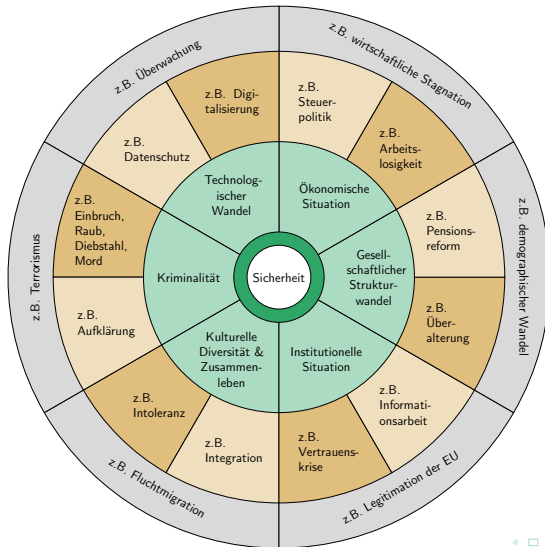
- FFG-finanziertes Projekt im Rahmen der KIRAS-Programmlinie
- Das Konsortium bestand aus:
 - Bedarfsträgern (BMLVS und BMI)
 - Wirtschaftsunternehmen (Repuco und RISE)
 - Wissenschaft (Universität Salzburg und Modul Universität Wien)
- Ziel war die Entwicklung einer Kooperationsplattform für die Bedarfsträger zur Entwicklung gemeinsamer Lagebilder für Österreich mit der Möglichkeit Trendanalysen und Szenariotechniken anzuwenden.

Rolle der Abteilung Soziologie im Projekt

Vier zentrale Herausforderungen:

1. Umfassende Konzeption von Sicherheit (Sicherheitsempfinden und gesellschaftliche Lageperzeption)
2. Bereitstellung einer möglichst vollständigen Liste von Medienquellen
3. Evaluation des Weblyzards
4. Entwicklung von Qualitätskriterien zur Bewertung von Medienquellen (Kooperation mit Abteilung Kommunikationswissenschaft)

1. Umfassende Konzeption von „Sicherheit“



2. Medienquellen – Funktionalität des Webyzards

Medienanalyse FORSTRAT PROJEKT

Hilfe | Abmelden

Suche #1

Dokumente 0-30 von 92096

Datumsauswahl 20 Jan 2017 - 20 Mar 2017

Quellen Mehrere

Interfaceoptionen Ungefiltert

Visualisierungen Export

TOPICS #2

Themen

Kriminalität	40205
Institutionelle Situation	24250
Kulturelle Diversität und Zus...	20362
Ökonomische Situation	15919
Gesellschaftlicher Strukturw...	12660
Technologischer Wandel	5070

Neue Kategorie hinzufügen

Anteil pro Thema #4

33%
20%
14%
11%
7%
10%

Wortwolke #6

Weltkarte #7

AKTUELLE SUCHE #3

92096

ASSOZIATIONEN

türkei	4511
polizei	11189
trump	5679
merkel	3596
straße	5250
wilders	1316
wien	8701
unfall	1853
spö	3531
erdogan	2909
feuerwehr	2336
unfall	2204
frauen	5797
schulz	1368
euro	14309

Prozess gegen Ex-HRE-Chef Funke beginnt #5

FP Prozess gegen Ex-HRE-Chef Funke beginnt. Die mögliche Höchststrafe für Funke sind drei Jahre Gefängnis. F. freipresse.de

Röttgen: Verteidigungsausgaben kein Wahlkampfthema

FP Röttgen: Verteidigungsausgaben kein Wahlkampfthema. Norbert Röttgen (CDU) spricht in der TV-Sendung ... freipresse.de

ORF eins: Glee

ORF eins: Glee - tv-ORF.at. Glee. Menschliche Requisite. Die Konkurrenz bei den 'Nationals' ist diesmal beson... tv.orf.at

ORF eins: Unforgettable

ORF eins: Unforgettable - tv-ORF.at. Unforgettable. Der letzte Check Out. Anwalt Brett Langley ist ermordet... tv.orf.at

ORF eins: Unforgettable

ORF eins: Unforgettable - tv-ORF.at. Unforgettable. In Flammen. Al und Carrie werden zu einem Tatort geruf... tv.orf.at

Chemnitz Jobcenter um 36.000 Euro betrogen

FP Chemnitz Jobcenter um 36.000 Euro betrogen. Foto: Frank Rumpenhorst/dpa-Archiv. Chemnitz/ Erfurt. W... freipresse.de

Busunglück: Gutachter schließt technische Mängel aus

FP Busunglück: Gutachter schließt technische Mängel aus. Ein Reisebus liegt bei Erfurt (Thüringen) umgekippt a... freipresse.de

Keine Beweise für Abhörung Trumps

FP Keine Beweise für Abhörung Trumps. US-Präsident Donald Trump geht in Washington, USA, zum Weißen Ha... freipresse.de

Neue Figur in "Sesamstraße" ist Autistin

ORF eins: Neue Figur in „Sesamstraße“ ist Autistin - news.ORF.at. Neue Figur in „Sesamstraße“ ist Autistin. Sie heißt Ju... orf.at

auto bad beamten bildung brand china deute #8

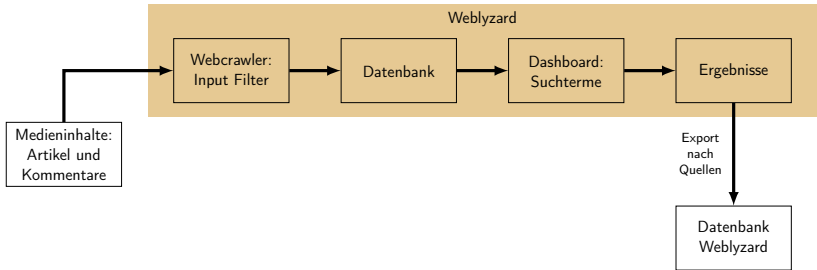
auto bad beamten bildung brand china deute deute#8 1
einsatzkräfte erdogan ermittler euro europa fahrer
fahrtbehinderung feuerwehr flüchtlinge fpo frauen gabriel
grünen informationen intervallen kinder kirchen koalition
kurz leiche leipziger linie merkel niederlande partei
pkw polizei politizien russland salzburg schulz soldaten
spö strafe too trump trumps täter türkei
türkischen unfall unterschiedlichen usa verkehrsunfall video
washington wien wilders zeugen övp

Netzwerkdiagramm #9

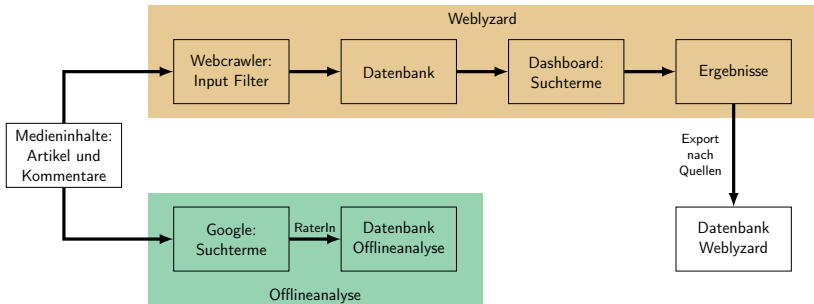
20 Mar 2017, 14:03 | FORSTRAT Projekt

powered by webyzard technology

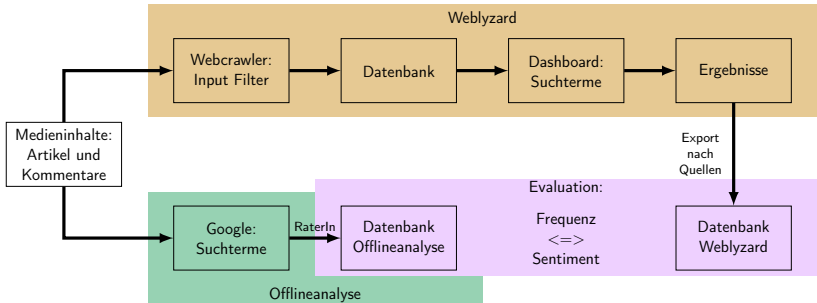
3. Evaluation des Weblyzard



3. Evaluation des Weblyzard



3. Evaluation des Weblyzard



Offlineanalyse I

- Ziel war ein Benchmark für das Stimmungsbild in Österreich.
- Qualitative Inhaltsanalyse kann diesen Benchmark am besten abbilden.
- Zeitraum, Quellen und Inhalte müssen im Weblyzard vorhanden sein:
 - 14-Tage-Zeitraum: 29.05.2017 bis 11.06.2017
 - linksliberales (Der Standard) und rechtskonservatives (Kronenzeitung) Medienspektrum
 - Drei Themenbereiche (Digitalisierung (Suchterm: Technologie), Arbeitsmarkt (Suchterm: Arbeitsmarkt), Radikalisierung (Suchterme: Radikalisierung, Rassismus, Muslime))

Offlineanalyse II

Die Artikel:

- Unabhängig von der Suchmaschine wurden alle Artikel verwendet.
- Zuordnung jedes Artikels zu einem Themenbereich und ggf. Kennzeichnung als Doppelnennung und irrelevante Artikel (Ausschluss aus weiterer Analyse)
- Der Standard: 138 (42) Artikel
- Kronenzeitung: 44 (11) Artikel
- Dreistufige Bewertung (positiv, neutral, negativ)

Offlineanalyse II

Die Artikel:

- Unabhängig von der Suchmaschine wurden alle Artikel verwendet.
- Zuordnung jedes Artikels zu einem Themenbereich und ggf. Kennzeichnung als Doppelnennung und irrelevante Artikel (Ausschluss aus weiterer Analyse)
- Der Standard: 138 (42) Artikel
- Kronenzeitung: 44 (11) Artikel
- Dreistufige Bewertung (positiv, neutral, negativ)

Die Kommentare:

- Kommentare wurden innerhalb eines Zeitfensters von 24 h bewertet (Höchstgrenze 200)
- Zahl der Kommentare ohne Höchstgrenze innerhalb des Evaluierungszeitraumes: 26.440
- Tatsächlich evaluierte Kommentare: 12.195
- Bewertung der Kommentare in positiv, negativ, neutral, zynisch, irrelevant
- Auch Antwort-Kommentare wurden klassifiziert und zählen zu der Obergrenze von 200 dazu

Inter-Rater-Reliabilität

- Inter-Rater-Reliabilität über die Artikel ($\kappa > 0,75$):
 - Kronenzeitung: $\kappa = 0,871$
 - Der Standard: $\kappa = 0,920$
- Korrelationen der BeurteilerInnen für die Kommentare nach Sentimentkategorie

	Positiv	Neutral	Negativ	Zynisch	Irrelevant	Mittelwert
Krone	0,966	0,745	0,980	0,927	0,721	0,868
Standard	0,782	0,969	0,977	0,949	0,858	0,907

Daten im Weblyzard

	Twitter-Accounts	Facebook	Twitter-Themen	Nachrichten	Kommentare
Arbeitsmarkt	17	18	1.812	35	11
Muslime	17	80	16.267	42	38
Radikalisierung	4	8	274	15	6
Rassismus	3	23	4.778	22	8
Technologie	4	14	251	61	4
Gesamt (Suche)	45	143	23.382	175	67
Gesamt (Zeitraum)	14.143	12.121	966.748	37.423 ^a	7.185 ^b
Suche/Zeitraum	0,003	0,012	0,024	0,005	0,009

^adavon 2021 aus dem Standard und 1931 aus der Kronenzeitung

^bdavon 4624 aus dem Standard und 986 aus der Kronenzeitung

- Extraktionsdatum: 12.07.2017

Dubletten

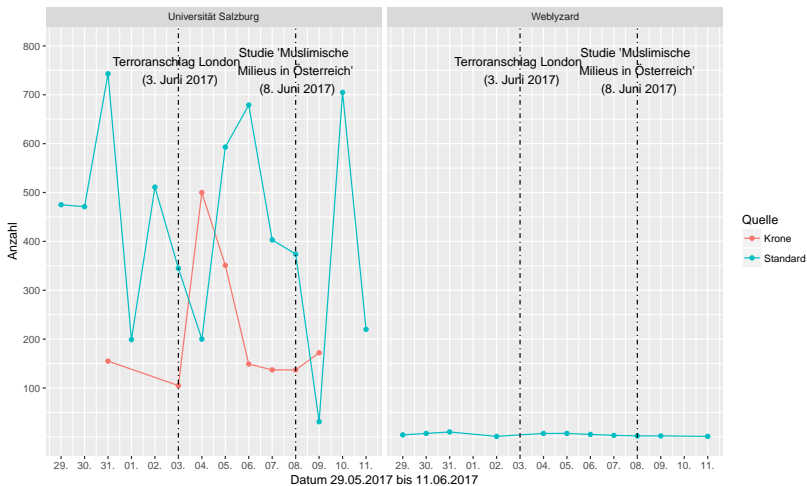
original_url	Datum	Titel	Sentiment	Relevance	Text
http://www.krone.at/videos/kern-will-buendnispartner-fuer-muslime-sein-gegen-terrorismus-video-573182?utm_source=krone.at&utm_medium=RSS-Feed&utm_campaign=%D6sterreich	08.06.2017	Kern will „Bündnispartner“ für Muslime sein	-0,71	2,98	bündnispartner, muslimen, islamistisch motivierten
http://www.krone.at/videos/digital/kern-will-buendnispartner-fuer-muslime-sein-gegen-terrorismus-video-573182	08.06.2017	Kern will „Bündnispartner“ für Muslime sein - Gegen Terrorismus - Video - Digital	-0,71	2,89	bündnispartner, muslimen, islamistisch motivierten
http://www.krone.at/nachrichten/kern-will-buendnispartner-fuer-muslime-sein-gegen-terrorismus-story-573109?utm_source=krone.at&utm_medium=RSS-Feed&utm_campaign=Nachrichten	07.06.2017	Kern will Bündnispartner für Muslime sein	-0,38	2,33	glaubensgemeinschaft, worte, zeichen
http://www.krone.at/oesterreich/kern-will-buendnispartner-fuer-muslime-sein-gegen-terrorismus-storydrucken-573109	07.06.2017	Kern will „Bündnispartner“ für Muslime sein - Gegen Terrorismus - Österreich	-0,61	2,57	glaubensgemeinschaft, kern, muna
http://www.krone.at/oesterreich/kern-will-buendnispartner-fuer-muslime-sein-gegen-terrorismus-story-573109	07.06.2017	Kern will „Bündnispartner“ für Muslime sein - Gegen Terrorismus - Österreich	-0,51	2,25	glaubensgemeinschaft, kern, duzdar

Frequenz der Kommentare

Suchbegriff	Offline		Facebook		Twitter		Kommentare	
Technologie	8,1%	(2160)	9,8%	(14)	1,1%	(251)	6,0%	(4)
Arbeitsmarkt	20,4%	(5398)	12,6%	(18)	7,7%	(1812)	16,4%	(11)
Radikalisierung	7,8%	(2053)	5,6%	(8)	1,2%	(274)	9,0%	(6)
Rassismus	11,2%	(2953)	16,1%	(23)	20,4%	(4778)	11,9%	(8)
Muslime	52,4%	(13876)	55,9%	(80)	69,6%	(16267)	56,7%	(38)
Summe	100%	(26440)	100%	(143)	100%	(23382)	100%	(67)

- Extrem wenig Kommentare im Weblyzard

Kommentare zu Radikalisierung im zeitl. Verlauf



Sentiment der Artikel

Tabelle 1: Sentimentverteilung aller Artikel

Sentiment	Uni Salzburg		Weblyzard	
positiv	6.0%	(13)	44.7%	(63)
neutral	34.3%	(74)	17.7%	(25)
negativ	38.9%	(84)	37.6%	(53)
uneindeutig	5.1%	(11)	0.0%	(0)
irrelevant	15.7%	(34)	0.0%	(0)

Tabelle 2: Sentimentverteilung der Artikel zu Radikalisierung

Sentiment	Uni Salzburg		Weblyzard	
positiv	2.0%	(2)	18.6%	(11)
neutral	26.7%	(27)	23.7%	(14)
negativ	55.4%	(56)	57.6%	(34)
uneindeutig	5.9%	(6)	0.0%	(0)
irrelevant	9.9%	(10)	0.0%	(0)

Sentiment der Kommentare

Tabelle 3: Sentimentverteilung aller analysierten Kommentare

Sentiment	Uni Salzburg	Weblyzard
positiv	4.7% (578)	37.5% (24)
neutral	33.3% (4061)	18.8% (12)
negativ	34.3% (4183)	43.8% (28)
zynisch	20.5% (2501)	0.0% (0)
irrelevant	7.2% (872)	0.0% (0)

Tabelle 4: Sentimentverteilung der analysierten Kommentare zu Radikalisierung

Sentiment	Uni Salzburg	Weblyzard
positiv	4.7% (362)	30.6% (15)
neutral	32.2% (2462)	18.4% (9)
negativ	36.3% (2775)	51.0% (25)
zynisch	20.5% (1571)	0.0% (0)
irrelevant	6.3% (485)	0.0% (0)

Fazit aus Evaluation

- Frequenz der Artikel weicht geringfügig ab (Dubletten!).

Fazit aus Evaluation

- Frequenz der Artikel weicht geringfügig ab (Dubletten!).
- Die Anzahl der Kommentare ist viel zu gering.

Fazit aus Evaluation

- Frequenz der Artikel weicht geringfügig ab (Dubletten!).
- Die Anzahl der Kommentare ist viel zu gering.
- Facebook und Twitter-Accounts generieren sehr wenige Daten.

Fazit aus Evaluation

- Frequenz der Artikel weicht geringfügig ab (Dubletten!).
- Die Anzahl der Kommentare ist viel zu gering.
- Facebook und Twitter-Accounts generieren sehr wenige Daten.
- Twitter-Themen ist eine brauchbare Ergänzung.

Fazit aus Evaluation

- Frequenz der Artikel weicht geringfügig ab (Dubletten!).
- Die Anzahl der Kommentare ist viel zu gering.
- Facebook und Twitter-Accounts generieren sehr wenige Daten.
- Twitter-Themen ist eine brauchbare Ergänzung.
- Die Funktion Nachrichtenmedien bildet den Diskurs gut ab.

Fazit aus Evaluation

- Frequenz der Artikel weicht geringfügig ab (Dubletten!).
- Die Anzahl der Kommentare ist viel zu gering.
- Facebook und Twitter-Accounts generieren sehr wenige Daten.
- Twitter-Themen ist eine brauchbare Ergänzung.
- Die Funktion Nachrichtenmedien bildet den Diskurs gut ab.
- Die Dynamik wird für die Nachrichten und Kommentare derzeit unzureichend erfasst.

Fazit aus Evaluation

- Frequenz der Artikel weicht geringfügig ab (Dubletten!).
- Die Anzahl der Kommentare ist viel zu gering.
- Facebook und Twitter-Accounts generieren sehr wenige Daten.
- Twitter-Themen ist eine brauchbare Ergänzung.
- Die Funktion Nachrichtenmedien bildet den Diskurs gut ab.
- Die Dynamik wird für die Nachrichten und Kommentare derzeit unzureichend erfasst.
- Das Sentiment im Weblyzard ist tendenziell positiver.

4. Entwicklung von Qualitätskriterien

Rating	Reichweite	Differenziertheit der Behandlung der Themen	Abbildung des Meinungsspektrums
●●●●●	Höchste Reichweite (umfassende Rezeption in der Bevölkerung)	Sehr hohe Aussagekraft/ Seriosität der Berichterstattung	breite Abdeckung der Meinungs- und Wertevielfalt in der Gesamtbevölkerung
●●●●○	Hohe Reichweite (hohe Auflage und breite Rezeption der Mediendiskurse)	Hohe inhaltliche Qualität der Berichterstattung, geringe Einschränkungen	Generalisierbare Schlussfolgerungen zu relevanten öffentlichen Diskursen durch aggregierte Datenanalyse möglich
●●●○○	Mittlere Reichweite (zumindest für einzelne Milieus in der Bevölkerung von Interesse)	Eingeschränkte Aussagekraft, Oberflächlichkeit in der Behandlung der Themen	Darstellung milieuspezifischer Diskurse (spezifische inhaltliche Ausrichtungen erkennbar)
●●○○○	Geringe Reichweite (für spezifisch interessierte NutzerInnen)	Deutlich eingeschränkte Aussagekraft, fragwürdige Art der Meinungsbildung	Einseitige Diskursdarstellung, erkennbar durch spezifisches NutzerInnenprofil
●○○○○	Äußerst geringe Reichweite („Nischenprodukte“, „hochspezialisiert“, „abseits vom Mainstream“)	Äußerst geringe Aussagekraft (Transport von Einzelmeinungen, Social Bots, Fake News)	Gefahr der Manipulation gegeben (z.B. durch „Filter Bubbles“ und „Echokammern“)

4. Bewertung von Medienquellen im Weblyzard

Kategorie	Reichweite	Differenziertheit der Themen	Abbildung des Meinungsspektrums
Nachrichten	● ● ● ● ○	● ● ● ● ○	● ● ● ● ○
Kommentare	● ○ ○ ○ ○	● ● ○ ○ ○	● ● ○ ○ ○
Facebook	● ● ● ○ ○	● ● ○ ○ ○	● ● ● ○ ○
Twitter-Accounts	● ● ○ ○ ○	● ● ○ ○ ○	● ● ○ ○ ○
Twitter	● ● ○ ○ ○	● ○ ○ ○ ○	● ● ○ ○ ○

Grundsätzliche Probleme bei Big-Data-Analysen

- Technische Unzulänglichkeiten

Grundsätzliche Probleme bei Big-Data-Analysen

- Technische Unzulänglichkeiten
- Zugänglichkeit der Daten (Eigentumsverhältnisse und Nutzungsmöglichkeiten)

Grundsätzliche Probleme bei Big-Data-Analysen

- Technische Unzulänglichkeiten
- Zugänglichkeit der Daten (Eigentumsverhältnisse und Nutzungsmöglichkeiten)
- Analysealgorithmen

Grundsätzliche Probleme bei Big-Data-Analysen

- Technische Unzulänglichkeiten
- Zugänglichkeit der Daten (Eigentumsverhältnisse und Nutzungsmöglichkeiten)
- Analysealgorithmen
 - Datenextraktion aus den Quellen (Schreibweise, Abkürzungen)

Grundsätzliche Probleme bei Big-Data-Analysen

- Technische Unzulänglichkeiten
- Zugänglichkeit der Daten (Eigentumsverhältnisse und Nutzungsmöglichkeiten)
- Analysealgorithmen
 - Datenextraktion aus den Quellen (Schreibweise, Abkürzungen)
 - Zuordnung von sarkastischen oder ironischen Postings

Grundsätzliche Probleme bei Big-Data-Analysen

- Technische Unzulänglichkeiten
- Zugänglichkeit der Daten (Eigentumsverhältnisse und Nutzungsmöglichkeiten)
- Analysealgorithmen
 - Datenextraktion aus den Quellen (Schreibweise, Abkürzungen)
 - Zuordnung von sarkastischen oder ironischen Postings
 - Fehlende Transparenz des Zuordnungsalgorithmus

Grundsätzliche Probleme bei Big-Data-Analysen

- Technische Unzulänglichkeiten
- Zugänglichkeit der Daten (Eigentumsverhältnisse und Nutzungsmöglichkeiten)
- Analysealgorithmen
 - Datenextraktion aus den Quellen (Schreibweise, Abkürzungen)
 - Zuordnung von sarkastischen oder ironischen Postings
 - Fehlende Transparenz des Zuordnungsalgorithmus
- Validität

Grundsätzliche Probleme bei Big-Data-Analysen

- Technische Unzulänglichkeiten
- Zugänglichkeit der Daten (Eigentumsverhältnisse und Nutzungsmöglichkeiten)
- Analysealgorithmen
 - Datenextraktion aus den Quellen (Schreibweise, Abkürzungen)
 - Zuordnung von sarkastischen oder ironischen Postings
 - Fehlende Transparenz des Zuordnungsalgorithmus
- Validität
- Umkehrung der Forschungslogik

Vielen Dank für Ihre Aufmerksamkeit!