

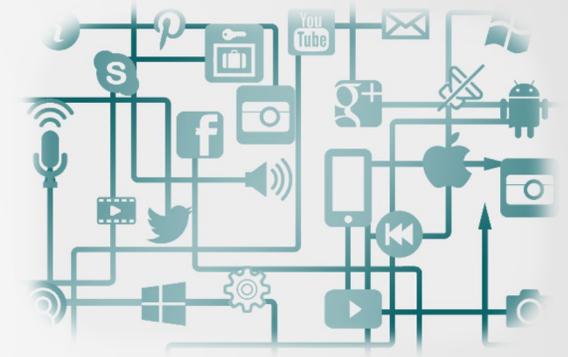
# How to handle unstructured textual data? A reflection on mixed-methods approaches for text analyses.

**Hein-Pensel, Franziska**

franziska.hein-pensel@tu-ilmenau.de

# Forschungsziel - *Neue Methodenkombinationen eröffnen neue Potenziale*

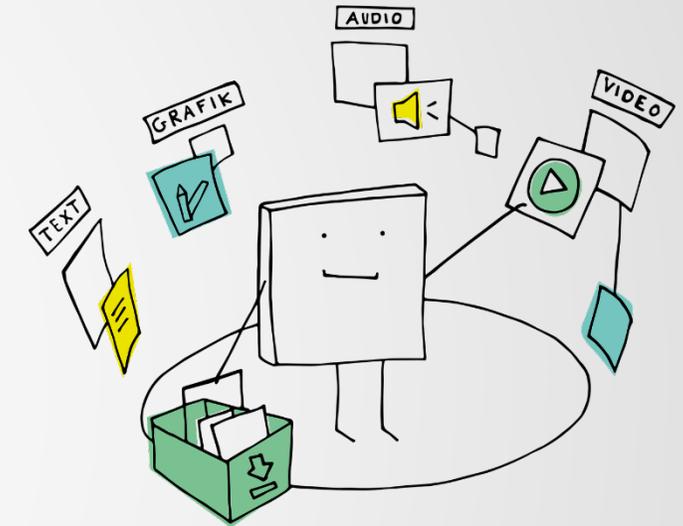
- *Datenmassen analysieren ohne Verzicht auf Einbettung in den Kontext*
  - quantitative Analyse*
  - qualitative Analyse*
- *Verborgene Struktur analysieren:*



- *Institutionelle und gesellschaftliche Mustererkennung*
- *Analyse branchenübergreifend bzw. über lange Zeiträume hinweg*
- *Verschiedene Datenquellen verknüpfen*

# Forschungsmotivation - *Wieso über Textanalyse in Mixed Method Ansätzen sprechen?*

- Quantitative computergestützte Textanalyse lange Historie
  - Erste Publikationen hierzu bereits 1966
- Wachstum an großen Datensätzen
  - Zunehmende Anzahl an Forscher nutzen quantitative Textanalyse
    - Häufig fehlende Transparenz
    - Kaum standardisierten Qualitätsrichtlinien



Preprocessing	Software			Total
	Not indicated	Non-proprietary	Proprietary	
Not indicated	19	9	28	56
Indicated	3	2	8	13
Detailed	13	21	21	55
Total	35	32	57	124

# Methodenkombination im Forschungsprozess

# 1. Forschungsfrage formulieren

Beschreibung	Reflexion
Klar spezifizierte Forschungsfrage/-richtung; Literatur überblicken	Was ist der derzeitige Forschungsstand? Welcher Datensatz steht mir zur Verfügung? ...



- Vorstellung Forschungsvorgehen an eigener Studie

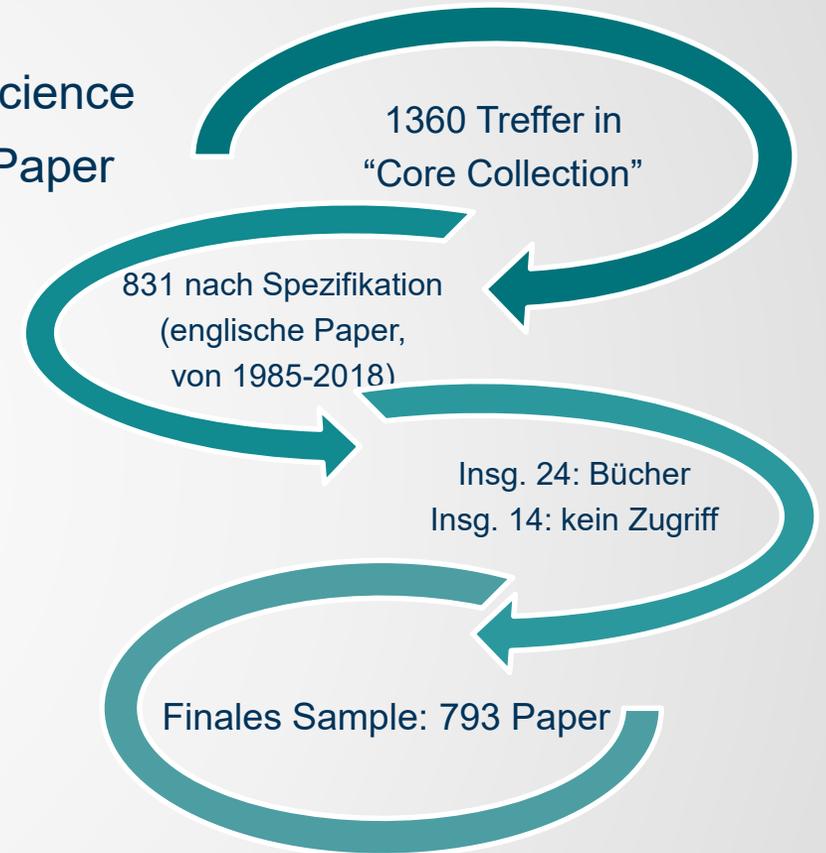
*Knorr, Karin; Hein-Pensel, Franziska: Since Albert and Whetten: The dissemination of Albert and Whetten's approach to characterize organizational identity, Accepted for the Crafting World Leading Outputs from Qualitative Research' conference, The University of Liverpool Management School, 2020. (Working Paper)*

- *Einbettung der Studie:* Organizationstheorien
- *Ziel:* Systematische Analyse der Verbreitung des Konzeptes organisationaler Identität von Albert und Whetten (1985)

## 2. Datensammlung

Beschreibung	Reflexion
<b>Interviews:</b> <ul style="list-style-type: none"> <li>• Forschungsgebiet kennenlernen</li> <li>• Vorgespräche (Fragen der Anonymität klären)</li> <li>• Interview-Protokoll</li> </ul>	<i>Sind die Daten geeignet?</i> <i>Welche Einschränkungen gibt es?</i> <i>Wie hat der Kontext die Interviewbedingungen und -ergebnisse beeinflusst?</i> ...
<b>Website:</b> <ul style="list-style-type: none"> <li>• Nutzungsbedingungen lesen</li> <li>• Manuelles oder automatisches Sammeln</li> <li>• Beschränkung auf eine Sprache zwecks Vergleichbarkeit</li> </ul>	<i>Sind die Daten geeignet?</i> <i>Welche Einschränkungen gibt es?</i> <i>Was wird angestrebt?</i> <i>Was wird nicht benötigt?</i> <i>Manuelle oder automatische Erfassung?</i> ...
...	

- *Datenbank:* Web of Science
- *Korpus:* vollständige Paper



# 3. Methodenauswahl

Beschreibung	Reflexion
Methoden auswählen, die am besten zum Forschungsziel und den Daten passen ➤ nicht nur was als "best practice" interpretiert wird	<i>Welche Bibliotheken stehen für die Sprache des Korpus zur Verfügung?</i> <i>Welche Quellen für Expertenunterstützung habe ich?</i>
Methoden kombinieren ➤ um so viele Erkenntnisse wie möglich über die Daten zu gewinnen	<i>Wie etabliert ist die Methode?</i> ...



Qualitative Kodierung

RQ1:

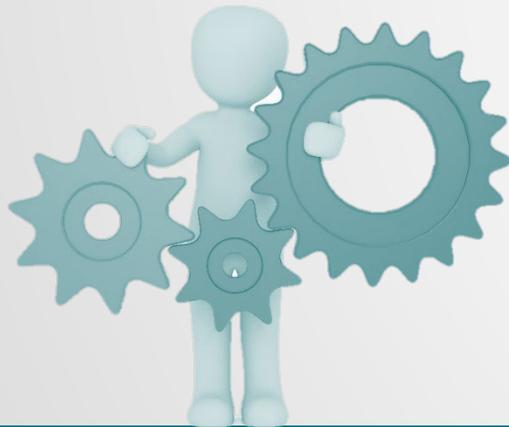
- a) *Which thematic scope on organizational identity referring to Albert and Whetten (1985) is shared by authors?*
- b) *Which elements embedded in their seminal paper disseminated into further research?*

RQ2:

- a) *Which elements, journals and author networks play a central role in adapting Albert and Whetten (1985) within management and business literature over time?*
- b) *Which elements, journals and author networks play a central role in adapting Albert and Whetten (1985) across fields outside of management and business literature over time?*

RQ3:

*Which aspects of Albert and Whetten's (1985) approach have been neglected in research so far and how can they complement future research?*



# 4. Databearbeitung für quantitative Analyse

Beschreibung	Reflexion
Mit Daten vertraut machen; Korpus definieren und konvertieren in eine Datenstruktur; Auf der Grundlage der Forschungsagenda Abstraktionsebene wählen	<i>Welche Informationen muss ich erhalten? Was ist meine Analyseeinheit? Welche Software oder welche Software-Tools sind am besten geeignet? Was ist die Einheit der Analyse? ...</i>

- **Konvertierung von pdf in txt**
- **Prüfen auf OCR-Fehlinterpretationen und manuelle bereinigen der Textdateien**
  - Extrahieren von Titel und Textkorpus
  - Löschen von Zeitschrifteninformationen, Referenzen, Danksagungen, Fußnoten und Tabellen/Abbildungen
- **Importieren der Daten in Python**
- **Datensatz vorbereiten**

(Entfernen von Stoppwörtern, Zahlen, Formeln; Korpus Lemmatisieren und Stemmen; hinzufügen von Metadaten, Kleinschreibung des Textkorpus)



# 5. Quantitative Textanalyse

Beschreibung	Reflexion
Visualisierung der Daten; Text in eine quantitative Matrix umwandeln (Beutel mit Wörtern, tf-idf, ...); Parameter definieren	<i>Wie werden die Daten verteilt? Wie viele Daten sind bei der Datenaufbereitung verloren gegangen? Sind die Dokumente thematisch kohärent oder weitgehend unstrukturiert? ...</i>



- **Quantitative Textanalyse: Topic Modeling**

- gensim Implementation für LDA
- “Optimale” Topic Number im Sinne der Kohärenz (k=14)
- Metadaten: Publikationsjahr, Forschungsfeld

- **Netzwerkanalyse**

- Vernetzt Forscher durch ihre Co-Autorenschaft

- **Keyword in context Analyse (KWIC)**

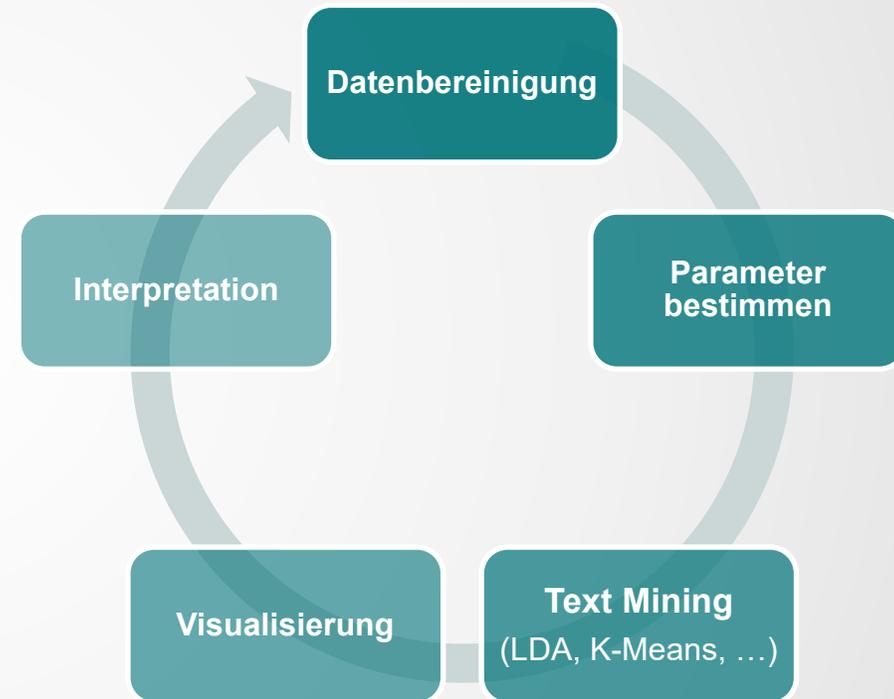
- Keyword String: "albert (and|&) whetten"
- Groß- und Kleinschreibung ignorieren

# 6. Qualitative Textanalyse – Teil I

Beschreibung	Reflexion
Ergebnisse überprüfen; gegebenenfalls Rückgriff auf Original Korpus; Aufgeschlossen bleiben und Analyse validieren	<i>Gibt es Fehler? Sind die Ergebnisse unterscheidbar (interpretierbar)? Ist die gewählte quantitative Methode für die Daten und Forschungsfrage angemessen? Können die Themen weiter geclustert werden? ...</i>

# 7. Zirkulieren

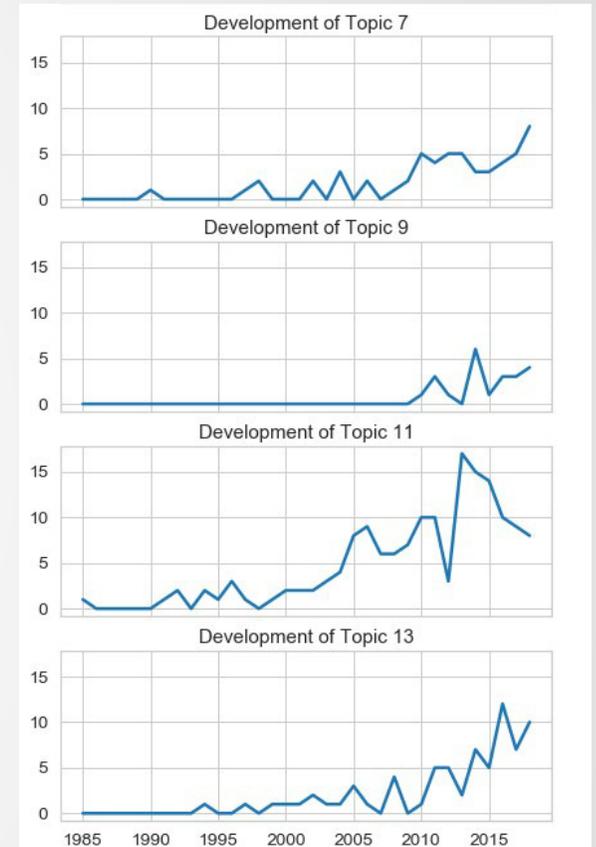
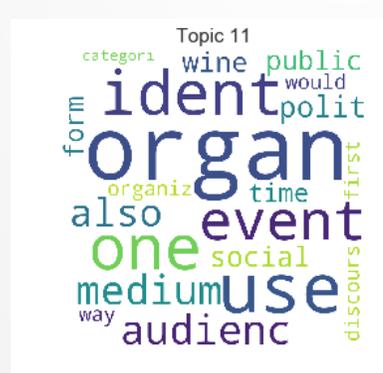
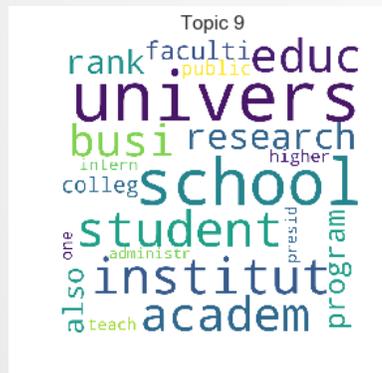
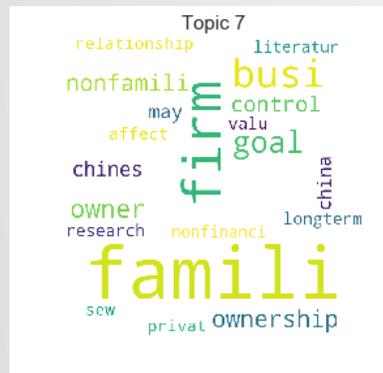
Beschreibung	Reflexion
Parameter verbessern; Fehlern beheben; Schritte dokumentieren ➤ Zwischen Schritt 5 und 6 zirkulieren	<i>Wann höre ich auf zu optimieren? Werden die Ergebnisse noch signifikant besser? ...</i>



# 8. Qualitative Textanalyse – Teil II

Beschreibung	Reflexion
Qualitativ Interpretation und Kodierung der Ergebnisse	<p><i>Was sind signifikante Ergebnisse?</i></p> <p><i>Wie können die Ergebnisse wieder mit dem Datensatz verbunden werden?</i></p> <p><i>Wie repräsentativ sind die Ergebnisse für meinen Datensatz?</i></p> <p>...</p>

- **Qualitative Textanalyse:**
  - Thematic label
  - Research stream
  - Abstract level



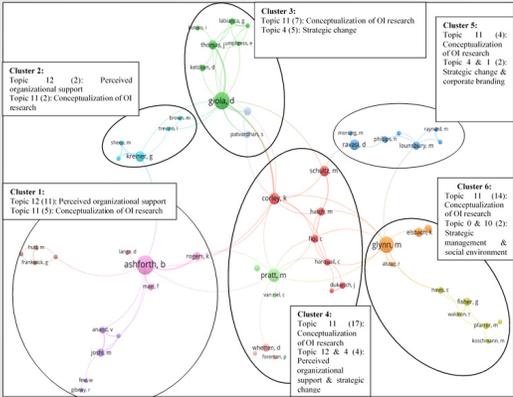
# 9. Ergebnisse zusammenfassen

Beschreibung	Reflexion
Visualisierung quantitativer Ergebnisse; Verknüpfung der Ergebnisse mit der Literatur; Methode transparent darstellen	<i>Welches Diagramm oder Schema beschreiben meine Ergebnisse am besten?</i> <i>Was sind die wissenschaftlichen Beiträge?</i> ...

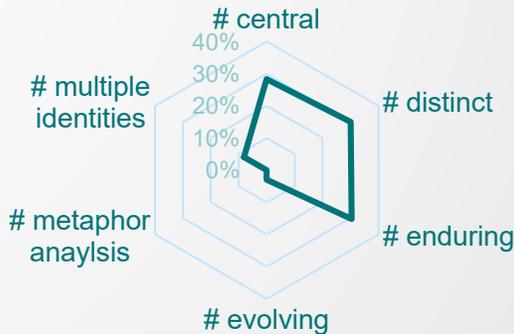
## Kodierschemata

Illustrative topics vocabularies	# Topic appearance	# Dominant documents (absolute size)	Impact sum/average	Mean topic probability	Highest probability	Main representatives	Research fields (Web of Science)	Thematic label	Research stream	Abstract level
Topic #00: firm, industri, perform, csr, stakehold, may, environment, market, sustain, use	396	6,81% (54)	4292/80	13,80%	72,80%	Decphouse, D.; Chen, Y.	Management; Business;	<ul style="list-style-type: none"> <li>Corporate social responsibilities (CSR)</li> <li>Organizational positioning</li> </ul>	Strategic management	Organizational Theory
...	...	...	...	...	...	...	...	<ul style="list-style-type: none"> <li>...</li> </ul>	...	...

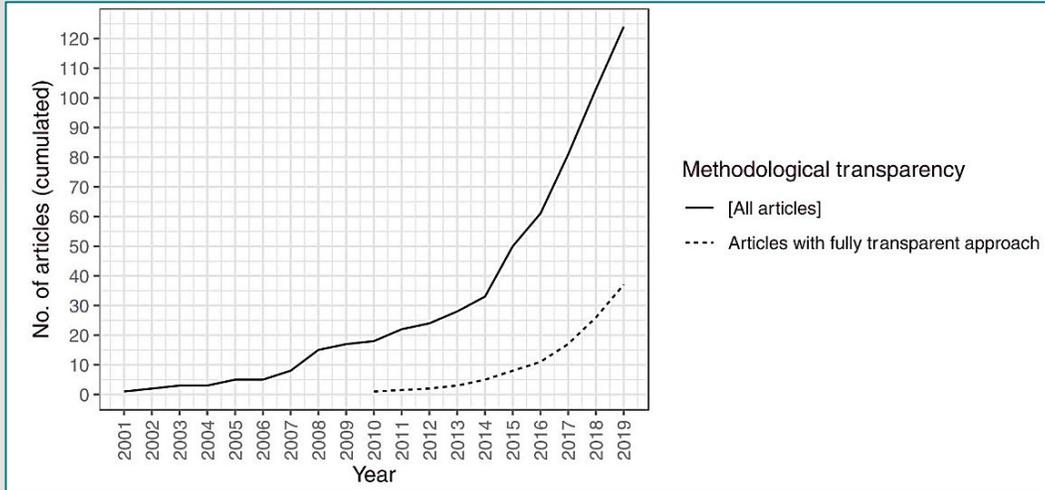
## Autorennetzwerke nach Themenfeldern



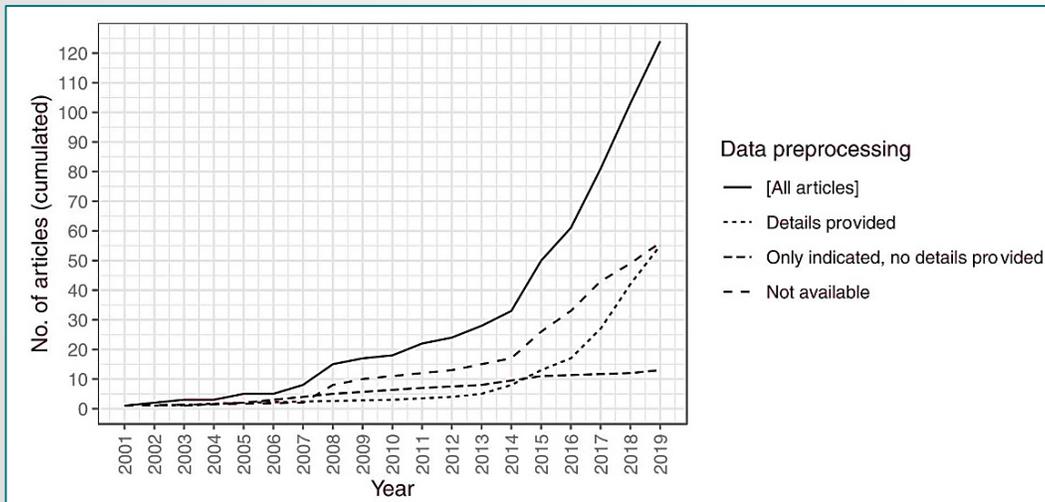
## Keyword in context analysis



# Ausblick



- Transparenz erhöht Sichtbarkeit und Attraktivität
- Kombination quantitativer und qualitativer Textanalyse
  - Repräsentatives Subsample ziehen
  - Analyse großer Datensätze möglich
  - Untersuchung gewinnt sowohl an Tiefe als auch an Breite



- Zusammenführung von Meta- mit Textdaten
  - Subjektive Faktoren werden einbezogen
  - Untersuchung komplexer sozialer Situationen
- Erkenntnisse stärken

**Vielen Dank!**