

# Das Projekt Automatische Sacherschließung an der ZBW

---

*Martin Toepfer (Wissenschaftlicher Mitarbeiter, M. Sc. – Informatik)*

*ZBW – Leibniz-Informationszentrum Wirtschaft*

*Workshop „Computerunterstützte Inhaltserschließung“ | UB Stuttgart, 8./9. Mai 2017*

---

# Vortragsinhalte

---

1. Die ZBW
2. Aktuelle Ausrichtung der automatischen Sacherschließung
3. Forschungsaspekte

# Kurzprofil: ZBW (2016)



- Das weltweit größte Informationszentrum für Wirtschaft.
- Bestand: ~4,43 Millionen Bücher und andere Veröffentlichungen.
- Anzahl der laufend gehaltenen Zeitschriften: 27.119 Titel.
- Rechercheportal: EconBiz, OA-Publikationsserver: EconStor, kontrolliertes Vokabular: Standard-Thesaurus-Wirtschaftswissenschaften (STW)
- Anzahl der Mitarbeiterinnen und Mitarbeiter: 280 Personen.
- Sitz: Kiel (Hauptsitz) und Hamburg,
- Finanzierung: Bund & Länder; Gesamtetat: 22,56 Millionen EUR

# Ziele

---

Was soll durch *automatische Sacherschließung* erreicht werden?

- Mehrwert für Nutzer in EconBiz (Suchportal, Information Retrieval), bspw.:
  - multilinguale Suche
  - Erschließungslücken überwinden
  - Themensuche mit dem hauseigenen Standard-Thesaurus Wirtschaft (STW)
- Homogene Sacherschließung unabhängig von der Höhe des Publikationsoutputs



# Herangehensweise

---

Vorgehen:

- keine Fokussierung auf elektronische Publikationen
- Ausrichtung der automatischen Erschließung an den jeweiligen Nutzungsrechten
  - Volltext
  - Metadaten (bspw.: Titel, Autorenkeywords)
- Anwendung/Re-Implementierung von Textmining-Verfahren in der ZBW-Umgebung
- Programmierung von Prototypen
  - webbasiertes Tool zur Bewertung automatisch generierter Indexate
  - neuartige Verfahren

# Forschungstransfer und Forschung in PB „A“

---

mit Forschungsabteilung „Knowledge Discovery“ der ZBW/CAU Kiel:

- Abschlussarbeiten und Praktika: bspw. Untersuchung verschiedener Systeme und Konfigurationen zur automatischen Indexierung, Thesaurusanreicherung

Aktuelle Forschung im Programmbereich „A“ und Kooperation mit Uni Passau

- „Fusion“-Ansatz: Kombination von sich ergänzenden Herangehensweisen
- Qualitätskontrolle, die kontrolliertes Vokabular (STW) und semantische Beziehungen der Begriffe berücksichtigt

# Übersicht: 2016/2017

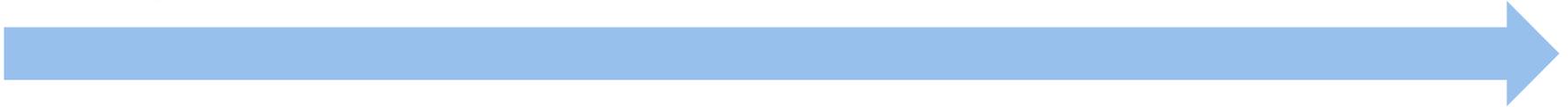
---

<= 2016:

verschiedene Projekte, erst mit  
kommerziellen Anbietern, später  
überwiegend studentische Arbeiten

2016/2017+ :

Fusionsansatz,  
differenzierte Qualitätskontrolle



2016:

initiale Arbeiten zu dokumentenorientierter Herangehensweise  
für Abbildung von Autorenkeywords  
Assoziativer Ansatz für Titel + Autorenkeywords

# Werkstattbericht: Titel + Autorenkeywords

---

z.T. sind Publikationen von den Autoren in eigenen Worten durch freie Schlagwörter beschrieben, die normiert werden sollen

- Crosskonkordanz-Ansatz: Abbildung einzelner Keywords auf STW-Begriffe („Nachnormierung“)
- „Dokumentenorientierter Ansatz“: nutze Kontext (zusammen auftretende Wörter des Dokuments); Integration mit Wörtern aus Titel
- assoziatives Lernen aus Beispielen: Zusammenhänge zwischen Wörtern und Begriffen

Ergebnisse:

- einige komplexe Verfahren, die zuvor gute Ergebnisse lieferten, sind zurückgefallen
- einfache Ansätze (mit weniger Parametern) lieferten vielversprechende Werte

Siehe Toepfer & Kempf (2016): <http://hdl.handle.net/11108/264>

---

# „Fusion“-Architekturen

---

kommender Beitrag auf der JCDL-2017

Kurzer Auszug:

- Analyse von Vor- und Nachteilen unterschiedlicher Herangehensweisen
- Kategorisierung der Ansätze
  - Lexikalisch
  - Statistisch, assoziativ
- abgestimmte Kombination von Verfahren
- Untersuchung bzgl. spezieller Anwendungsszenarien

# Fazit

---

- ZBW treibt automatische Indexierung voran & arbeitet an
  - neuen Herangehensweisen und Verfahren,
  - differenzierter Qualitätsmessung und –sicherstellung.

Martin Toepfer (Wissenschaftlicher Mitarbeiter, M. Sc. – Informatik)  
*ZBW – Deutsche Zentralbibliothek für Wirtschaftswissenschaften*  
*Leibniz-Informationszentrum Wirtschaft*

Düsternbrooker Weg 120  
24105 Kiel  
T: +49 431 8814-332  
E: m.toepfer@zbw.eu



Leibniz-Informationszentrum  
Wirtschaft  
Leibniz Information Centre  
for Economics

