

„Qualitätsgesteuerte“ maschinelle Inhaltserschließung wirtschaftswissenschaftlicher Publikationen

Tobias Reholz & Martin Toepfer
ZBW – Leibniz-Informationszentrum Wirtschaft

Workshop „Erschließung per Knopfdruck? Ein Spagat zwischen Quantität und Qualität?“
| 1. Juni 2017 | 106. Deutscher Bibliothekartag, Frankfurt am Main

Automatische Sacherschließung in der ZBW

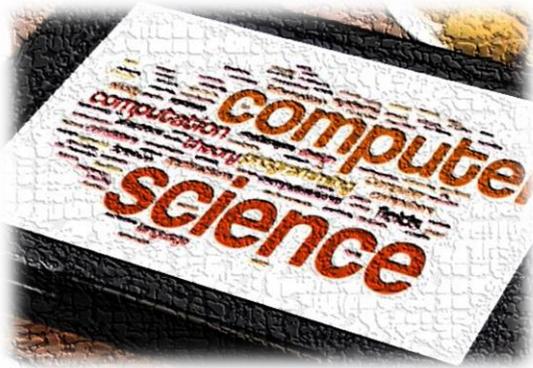
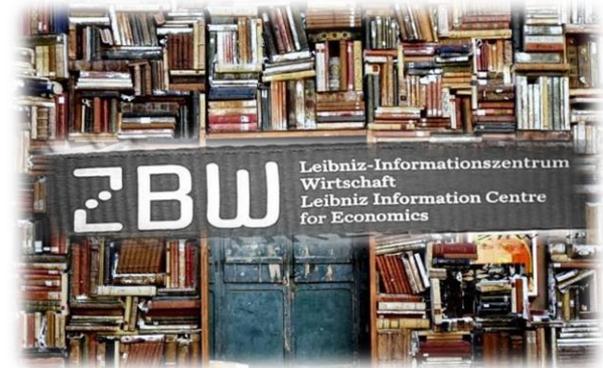


Photo: Nick Youngson CC BY-SA 3.0- <http://nyphotographic.com/>



- ✓ Anwendungsorientierte Forschung
- ✓ **Forschungsbasierte Eigenentwicklungen** im Bereich der maschinellen Erschließung
 - ✓ Interdisziplinäres Team ► direkter Austausch und **gemeinsame Ideenentwicklung**
- ✓ **Transparenz und Akzeptanz** ► Einbindung des Fachreferats in alle Entwicklungsprozesse

Quantität, Qualität...und unser Weg

BEST QUALITY

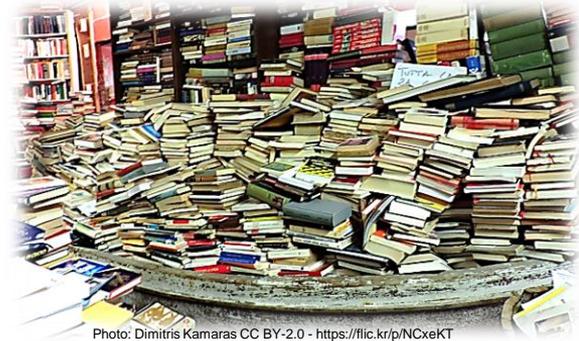


Photo: Dimitris Kamaras CC BY-2.0 - <https://flic.kr/p/NCxeKT>

- ✓ Erschließung auf Basis von inhaltlichen Metadaten ► **Print/Online** – jenseits rechtlicher Einschränkungen
 - ✓ **Qualität geht vor** ► Inhaltliche Normierung mit dem Ziel einer hohen Precision
- ✓ Kontinuierliche Weiterentwicklung der Verfahren zur Erhöhung des Recalls ► „**Dynamisches Indexat**“
„**Doppelter Boden**“ ► Globale Regel: mindestens zwei Verfahren müssen Deskriptor vorschlagen

„Fusion“-Architekturen

black-box?

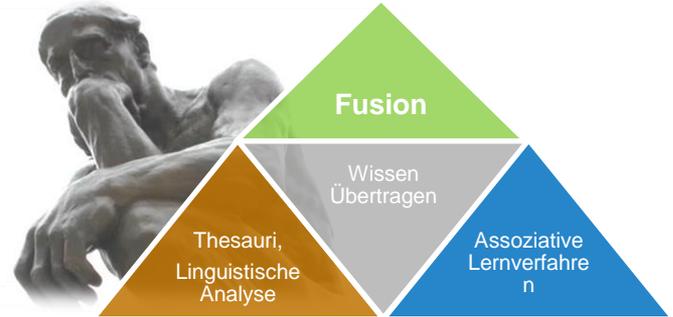


Photo: Eugene Kim CC BY-2.0 - <https://flic.kr/p/cYvZDC>

- ✓ „**Methoden verstehen**“ ► Analyse von Vor- und Nachteilen der Erschließungsmethoden
- ✓ „Best of all worlds“ ► Vorteile der einzelnen **Methoden** herausstellen und **gezielt weiterentwickeln**
 - ✓ Prototyp „*ZBW Metadata Indexing*“ ► abgestimmte **Kombination von Verfahren**
- ✓ **Plausibilitätsprüfung** ► Thesaurus-Systematik und semantische Beziehungen der Begriffe werden berücksichtigt

Bewertung maschineller Verfahren

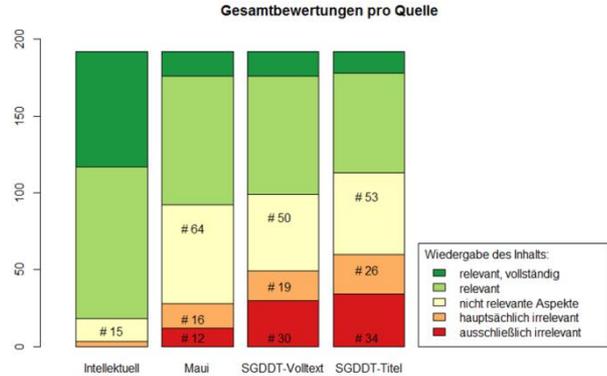


Photo: flickr: Tim Reckmann - CC BY-NC-SA 2.0

- ✓ Ziel: „**faire**“ **Bewertung** der maschinellen Verfahren ► Qualitätsbewertung jenseits von „Goldstandards“
- ✓ „Qualität“ der intellektuellen Erschließung als Benchmark ► **Bestimmung der Interindexierungskonsistenz**
 - ✓ **Intellektuelle Bewertung** ausgewählter Verfahren ► Ergänzung zu maschineller Auswertung
- ✓ Differenzierte Betrachtung der vorgeschlagenen Deskriptoren ► **Fehleranalyse zur Verfahrensverbesserung**

„Release-Tool“: Qualitätssicherung

1: Intellektuelle
Erschließung des
Dokuments

2: Bewertung
einzelner Schlagwörter

3: Bewertung
insgesamt

ZBW-A1-Release-Tool^{prototype} Home Stats Collection - prev next Help About

Econbiz-ID: 10002772736

Title: Differences in desired and actual fertility : an economic analysis of the Spanish case

Keywords: fertility religion family planning desired number of children unemployment

Abstract: Family size is the outcome of sequential decisions influenced both by preferences and by ongoing changes in the environment where a family lives. During the last two decades the gap between the number of children women prefer and their actual fertility has widened in Spain. The paper uses the 1985 and 1999 Spanish Fertility Surveys to study whether the tightening of the labor market and worsening of economic conditions in Spain during the last twenty years are important determinants of this change. I find that women facing high unemployment rates in their mid-twenties tend to restrict their fertility below their ideal level. Among working-women, the stability of a public sector job lessens the difficulties of balancing work and family and of achieving preferred fertility. Temporary contracts work in the opposite direction. Findings are robust to the inclusion of controls for the use of family planning as well as within-couple discrepancies in either preferences or religious affiliation. -- fertility ; desired number of children ; unemployment ; family planning ; religion

Descriptor	main_ai	lazy	maui	binary_relevance
15939-4 Family planning V N	👇	👇	👇	
15941-3 Fertility N	👇	👇	👇	
15925-1 Children N	👇	👇	👇	
15671-6 Religion N	👇	👇	👇	
15097-5 Index construction N				
13427-6 Fertilizer P	👇			
17565-2 China G				
17161-5 Spain G				

Automatisch erstellter **Hauptvorschlag** unter Einbezug der Thesaurussystematik

Erklärungskomponente für unterschiedliche Verfahren:

- wörterbuchbasiert
- statistisch, assoziativ
- nächste Nachbarn
- ...

Total rating: Reject Accept, Fair Accept, Good

Fazit

Die ZBW treibt die automatische Indexierung voran & arbeitet an

- neuen Herangehensweisen und Verfahren,
- differenzierter Qualitätsmessung und –sicherstellung.



Photo: flickr: Esko Kurvinen - CC BY-NC 2.0

Tobias Rebholz, Martin Toepfer
ZBW – Deutsche Zentralbibliothek für Wirtschaftswissenschaften
Leibniz-Informationszentrum Wirtschaft

Düsternbrooker Weg 120
24105 Kiel
E: t.rebholz@zbw.eu
E: m.toepfer@zbw.eu