

Automatische Sacherschließung an der ZBW – Status quo & Ausblick

Thomas Groß, MA (LIS)

Deutsche Zentralbibliothek für Wirtschaftswissenschaften
Abteilung Informationsdienste

**Jahrestagung der Gesellschaft für Klassifikation -
Bibliothekarische Fortbildung/ Workshop**

Frankfurt am Main, 31.08.2011

Übersicht

1. Einführung – Grundsätzliches
 2. Projektphasen – Rückblick und Ergebnisse
 3. Mögliche Szenarien – Ausblick
- Weiterführende Literatur

1. Einführung – Grundsätzliches I

→ Allgemein:

- automatisierte Sacherschließung (Indexierung, Klassifikation) elektronischer und später aller Informationsressourcen

→ Konkret:

- Auswahl, Prüfung und Evaluierung maschineller Verfahren
- organisatorische Einbindung und begleitende Evaluierung

→ Standard-Thesaurus Wirtschaft (STW):

- ca. 6.200 Deskriptoren, dichte semantische Relation
- polyhierarchisch und bilingual (dt./engl.)

→ Input

- ca. 15.000 - 20.000 Online-Working-Paper
 - Online-Diss., (parallele) Online-Zeitschriftenaufsätze
-

1. Einführung – Grundsätzliches II

- Projektteam (unterschiedliche Zeitanteile):
 - Manfred Faden (Fachreferent, HH): Leitung
 - Thomas Groß (Fachreferent, KI): u. a. Auswertung
 - Max-Michael Wannags (Dipl.-Inf., HH): IT-Aspekte

- offen: (nicht Projektauftrag)
 - Einbindung IT-Entwicklung/ Forschung!?
 - Input (Formalerschließung)
 - Output (Zielsystem: GBV/ EconBiz)

2. Projektphasen I

→ Probeläufe: Mitte – Ende 2009

- Test von zwei Verfahren (statistisch vs. linguistisch)
- Kaufentscheidung: Decisiv Categorization
 - statistischer Ansatz + STW = begriffsorientiertes Verfahren



→ wesentliche Ergebnisse: Anfang 2010

- 1/3 Indexierungskonsistenz Mensch/Maschine
- STW-Nutzung durch ReferentInnen umfassender und trennschärfer (= fehlende/schmale Trainingsbasis)
- STW-Polyhierarchie als Problem
- deutsch (-) vs. englisch (+), VWL (+) vs. BWL (-),
themenspezifisch: z. B. Geldpolitik/Finanzmarkt (+)

2. Projektphasen II

- Implementierung (EDV-seitig): Anfang 2010
 - Installation & Schulung der Content-Administratoren (M. Faden/ T. Groß)

- Update auf Decisiv 7 & Hosting bei Recommind: Herbst 2010
 - fehlendes VPN, neues/komplexeres Admin-Tool
 - Software-Probleme, Anpassung durch Recommind (bis Frühjahr 2011)
 - interne Content-Administration schwierig → Outsourcing an Recommind

2. Projektphasen III

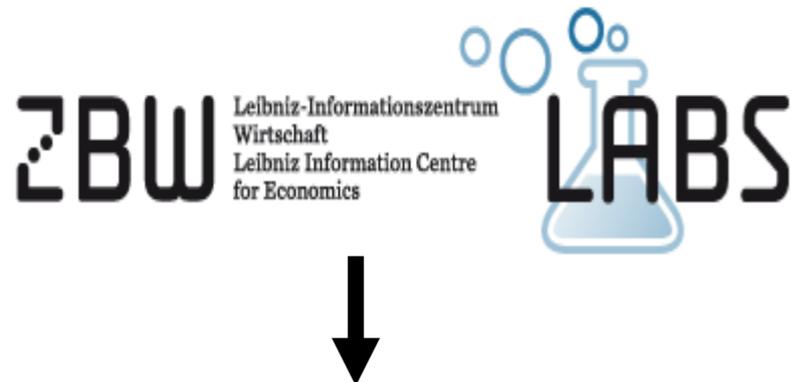
→ Erweiterung der Trainingsbasis: seit Mitte 2010

- Econbiz-Abzug, Verlags-Verhandlungen (gescheitert)
- aktuell ca. 200.000 Doks (davon ca. 1/3 indexiert)
- pdf-Problematik (Archivierungsstandards)

→ STW-Anpassung:

- flacher Thesaurus
- Systematik (Klassifikation): BWL, VWL, Geo's, Sektoren
- Sperrung problematischer Deskriptoren (Bsp. Atomwaffe, Zweitwohnung)

2. Projektphasen IV – Der ZBW- Indexer I



ZBW-Indexer

- „ZBW-Werkstatt“
 - innovative Anwendungen/ Services
 - Prototypen, Beta-Version
-
- Basis: Decisiv Categorization (Recommind), 83.297 Datensätze, STW
 - Funktion: Generierung potentieller STW-Schlagwörter aus wirtschaftswissenschaftlichen Texten
 - Ziel: auch Klassifikation

2. Projektphasen IV – Der ZBW- Indexer II

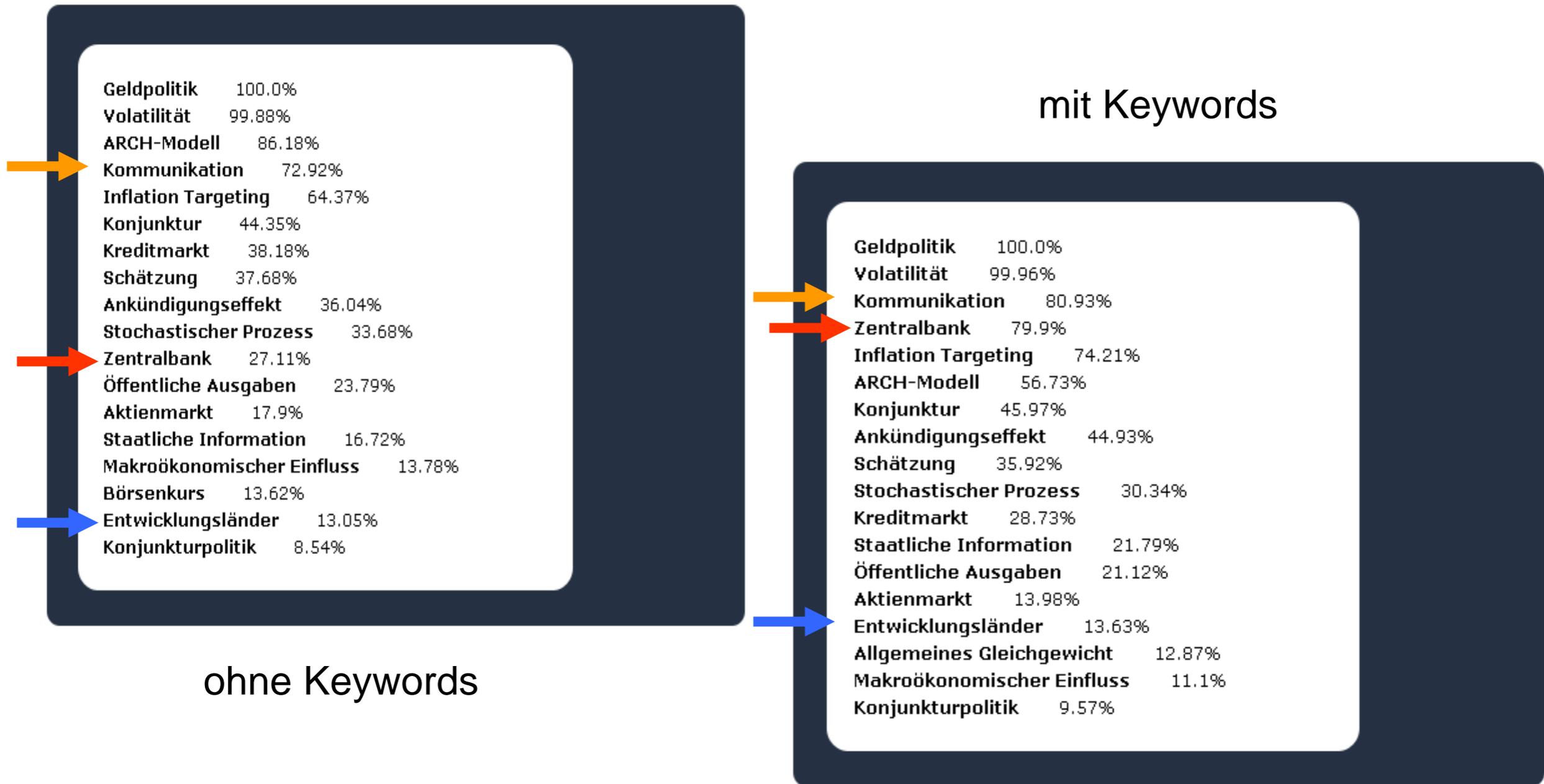
<http://zbw.eu/beta/zbw-indexer>

- Probetext:

Using a GARCH model, we analyze the influence of U.S. monetary policy action and communication on the price volatility of commodities for the period 1998-2009. We find, first, that U.S. monetary policy events have an economically significant impact on price volatility. Second, expected target rate changes and communications decrease volatility, whereas target rate surprises and unorthodox monetary policy measures increase it. Third, we find a change in reaction to central bank communication during the recent financial crisis: the calming effect of communication found for the whole sample is partly offset during that period. -- **Central Bank**

Communication ; Commodities ; Federal Reserve Bank ; Monetary Policy ; Price Volatility

2. Projektphasen IV – Der ZBW- Indexer III



3. Mögliche Szenarien I

→ Szenario 1: Klassifikation/Systematik

- jederzeit möglich, d. h. „startklar“
- Erschließungsinstrument:
 - STW-Systematik: rund 300 Stellen
 - JEL-Klassifikation???
- Training per Hand (Tax-Brower), anschließende Evaluierung
- grobe Sacherschließung auch ZBW-fremder Ressourcen möglich (EconBiz als Zielsystem)
- aufwendig, ca. 1 Fachrefent/in p. a. (VZÄ) nötig = keine unmittelbare Wirkung auf Indexierungslast!

3. Mögliche Szenarien II

- **Szenario 2: semiautomatische Indexierung im Regelbetrieb**
- Erschließungsinstrument: STW (flache Hierarchie)
 - Einarbeitung der SacherschließerInnen und begleitendes Training der Software durch tägliche Indexierung
 - neue Geschäftsabläufe (Formalerschließung, Zielsysteme)
 - konkreter, aber mittelfristiger Nutzen: Entlastung der Indexierung
 - sehr aufwendig: andere Abteilungen einbinden, zusätzlicher Einarbeitungsaufwand intern (Abteilung Info-Dienste)
 - Beispiele zeigen mehrjährige Implementierungsdauer

3. Mögliche Szenarien III

→ **Szenario 3: automatische Indexierung als Forschungsfeld**

- Haben: Aufbau einer Forschungsabteilung gestartet, Reorganisation läuft an
- Soll: begleitende, software/IT-gestützte Evaluierung (i. S. v. Indexierungskonsistenzen, auch Retrievaltests (recall/precision))
- nicht vorgesehen (abteilungsintern nicht möglich)

Vielen Dank für Ihre Aufmerksamkeit

Thomas Groß: t.gross@zbw.eu

Weiterführende Literatur

Groß, Thomas; Manfred Faden (2010): Automatische Indexierung elektronischer Dokumente an der ZBW. In: Bibliotheksdienst, Bd. 44 (2010), Heft 12, S. 1120-1135. [Link](#)

Groß, Thomas (2010): Automatische Indexierung von wirtschaftswissenschaftlichen Dokumenten: Implementierung und Evaluierung am Beispiel der ZBW. Berliner Handr., Heft 284. [Link](#)

Vorträge (Faden/Groß):

Petrus-Workshop, DNB, FFM, 21./22. März 2011, [Link](#)

14. GBV-Verbundkonferenz, Berlin, 08.09.2010, [Link](#)

FIS-Bildung-Tagung, Frankfurt am Main, 03.-04.05.2010, [Link](#)