Automatische Sacherschließung an der ZBW – Status quo & Ausblick

Manfred Faden (Dipl. Soz-Ök.; MA LIS,
Thomas Groß (Dipl. Pol.; MA LIS)
Deutsche Zentralbibliothek für Wirtschaftswissenschaften
Abteilung Informationsdienste
Petrus-Workshop an der Deutschen Nationalbibliothek
Frankfurt am Main
22.03.2011

Übersicht

1. Einführung

→ Allgemein:

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

→ Konkret:

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

→ Projektphasen:

- Probeläufe (2009)
- Update auf Decisiv 7 & Hosting bei Recommind (Herbst 2010)
- Implementierung (2010-)

2. Rückblick I

- Standard-Thesaurus-Wirtschaft (STW)
  - ca. 6.200 Deskriptoren, dichte semantische Relation

- Automatisches Indexierungsverfahren 

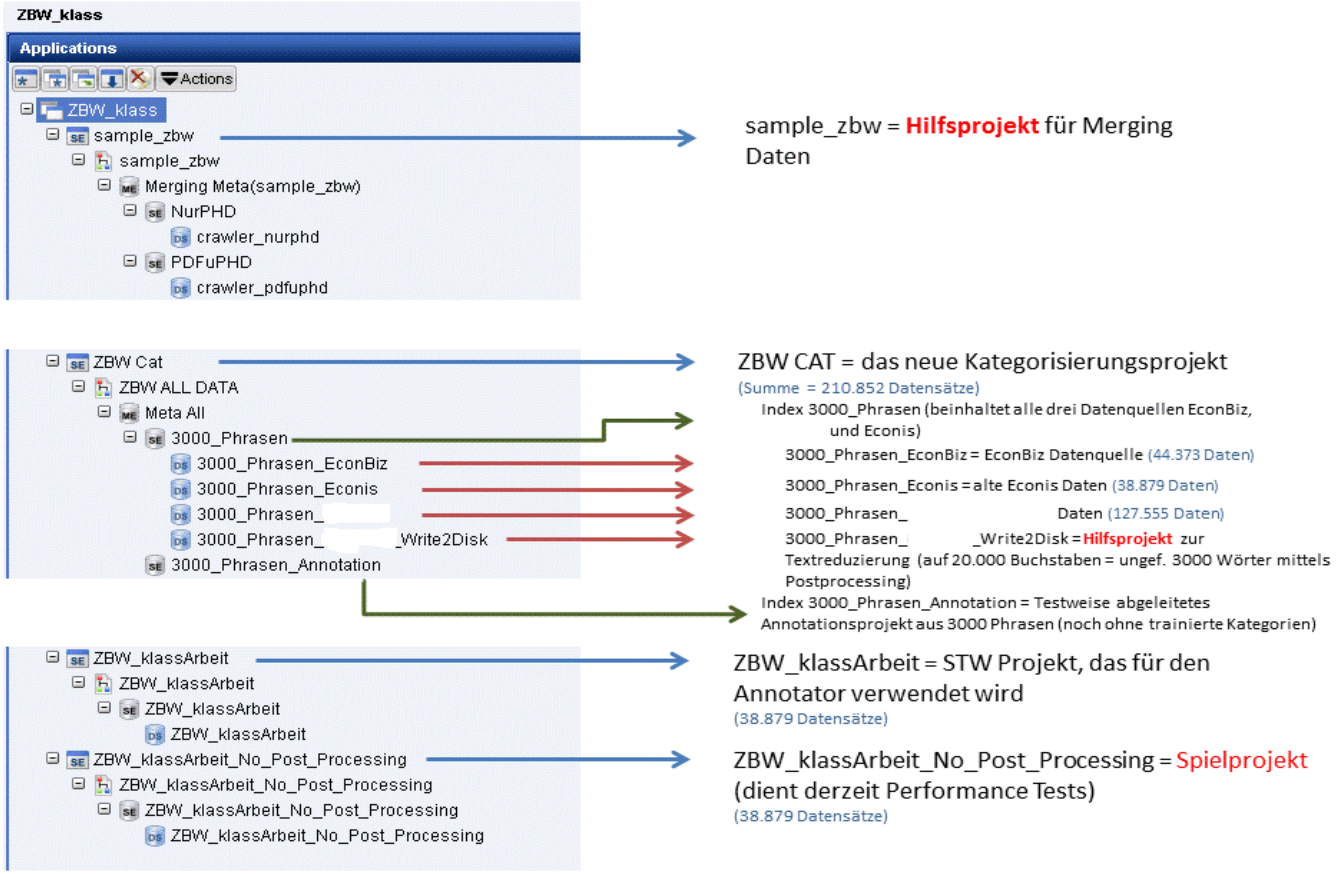
  - Decisiv Categorization
  - statistischer Ansatz + STW = begriffsorientiertes Verfahren

2. Rückblick II

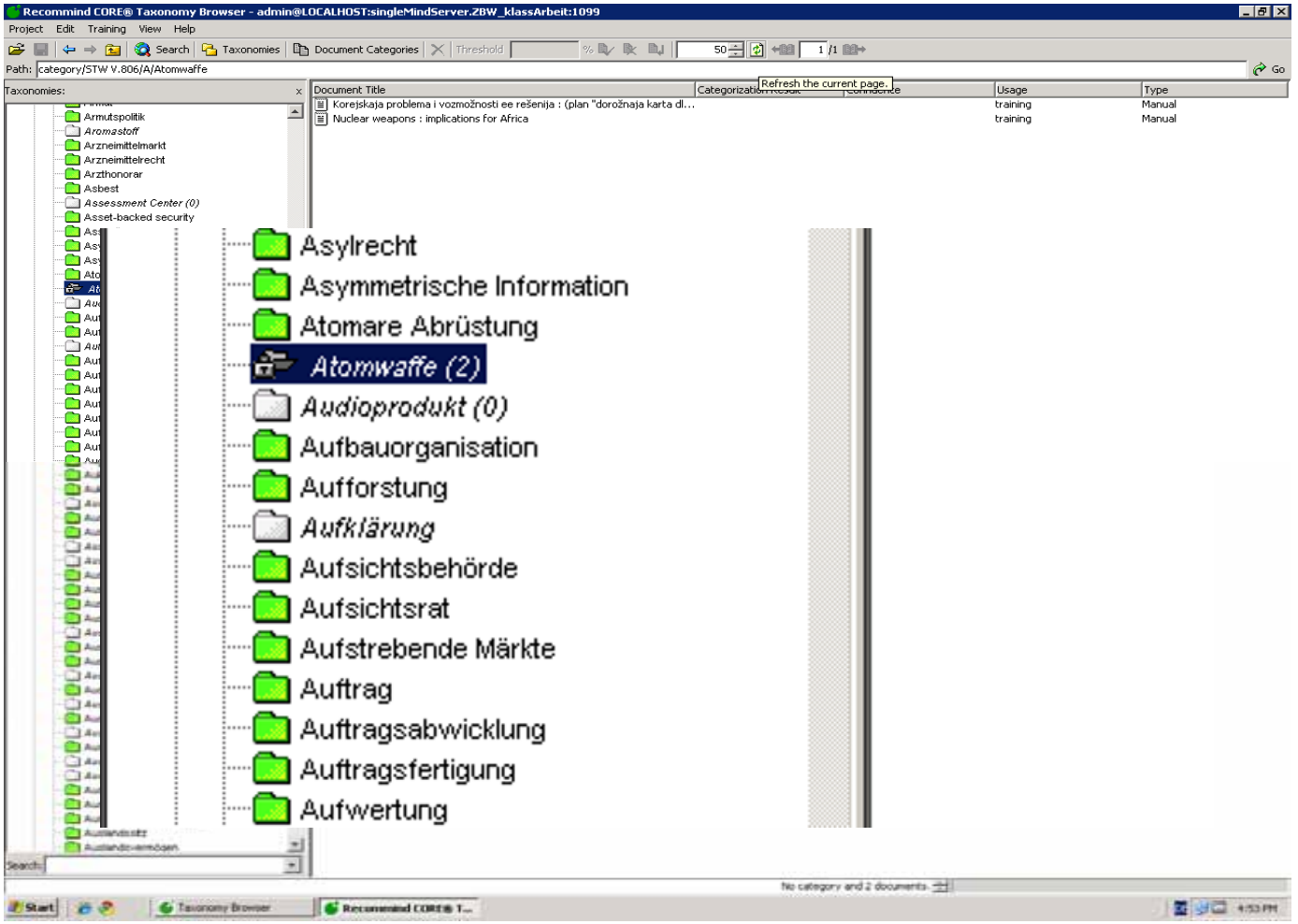
- Probeläufe/Implementierung (2009/2010)
  - Probleme:
    - Datenmaterial (broken links)
    - Trainingsbasis (2/3 nur indexiert, Tendenz abnehmend)
    - STW-Hierarchie (polyhierarchisch vs. flach)

  - Ergebnisse
    - 1/3 Indexierungskonsistenz
    - intellektuelle Indexierung trennschärfer, umfassendere STW-Nutzung

3. Laufende Projekte – Grundsätzliches I



3. Laufende Projekte – Grundsätzliches II



3. Laufende Projekte – Der ZBW- Indexer I

- „ZBW-Werkstatt“
- innovative Anwendungen/ Services
- Prototypen, Beta-Version

ZBW-Indexer

- Basis: Decisiv Categorization (Recommind), 83.297 Datensätze, STW
- Funktion: Generierung potentieller STW-Schlagwörter aus wirtschaftswissenschaftlichen Texten
- Ziel: auch Klassifikation

3. Laufende Projekte – Der ZBW- Indexer II

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

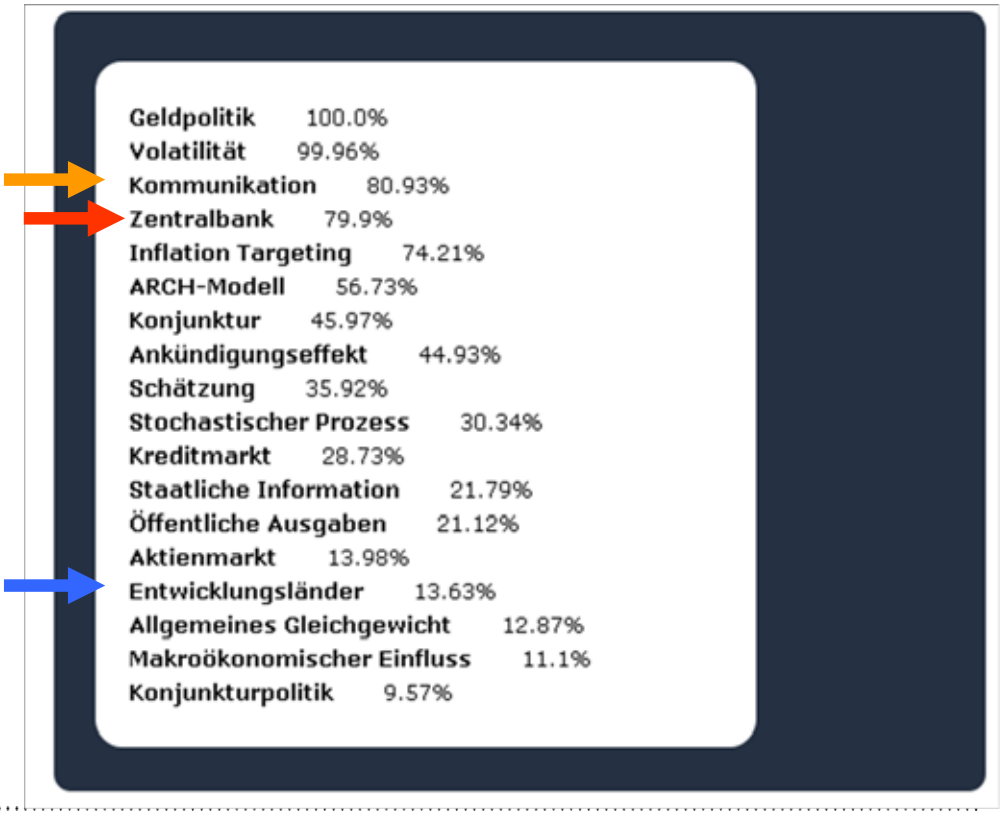
- Probetext:

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***

3. Laufende Projekte – Der ZBW- Indexer III

ohne Keywords

mit Keywords



3. Laufende Projekte – Der ZBW- Indexer IV

- Ergebnisse:
  - deutsch (-) vs. englisch (+)
  - Abstracts/Keywords (+) vs. lange Texte (-)
  - formellastig (-) vs. textlastig/Empirie (+)
  - VWL (+) vs. BWL (-)
  - themenspezifisch: z. B. Geldpolitik/Finanzmarkt (+)

3. Laufende Projekte – Trainingsbasis erweitern

- Trainingsbasis momentan zu klein, um viele Schlagwörter zu trainieren
- Verhandlungen (mit einem Verlag) über Datenlieferungen
  - elektronische Versionen von Artikeln, die von der ZBW in den vergangenen Jahren anhand der gedruckten Version schon erschlossen und in der Datenbank abgelegt worden sind

4. To Do: 2011

- → Erweiterung der Trainingsbasis
- → semiautomatische Indexierung/Klassifizierung (Test)
- → Geschäftsabläufe (Formalerschließung, Zielsysteme)
- → Klassifizierung: Implementierung/Evaluierung
- → Neu: Semtinel (Kai Eckert, UB Mannheim) als Evaluierungstool? (siehe: http://semtinel.org)

Vielen Dank für Ihre Aufmerksamkeit

Manfred Faden: m.faden@zbw.eu
Thomas Groß: t.gross@zbw.eu

Anhang

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 Handreichungen, Heft 284. Link

Vorträge:

14. GBV-Verbundkonferenz, Berlin, 08.09.2010, Link

FIS-Bildung-Tagung, Frankfurt am Main, 03.-04.05.2010, Link

