



Entwicklung eines Werkzeugs zur Visualisierung der SWD/GND

Dr.-Ing. Jan F. Maas

Hildesheim, den 1. August 2012

Überblick

1. Motivation für die Visualisierung SWD/GND
2. Eigenschaften / Programmdemonstration
3. Perspektive: GND

Visualisierung: Anwendungsszenarien

Anwendungsgebiet Fachreferat

- Unterstützung der Verschlagwortung von Medien
- Unterstützung der Ansetzung
- Unterstützung der Pflege der Normdatei(en)

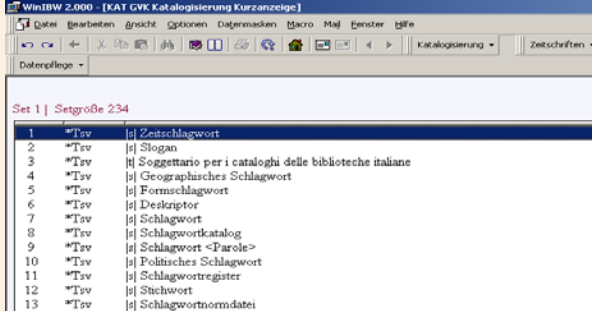
Anwendungsgebiet Recherche?

- Unterstützung der Recherche durch Navigation entlang der Hierarchien

Anwendungsgebiet: Verschlagwortung

Zugriff auf die SWD im Fachreferat (WinIBW)

- Trunkierte/exakte Suche von Schlagwörtern in WinIBW (Haupt- und Nebenansetzungsformen)
- Darstellung einzelner Datensätze im PICA3-Format



WinIBW 2.000 [KAT CVK Katalogisierung Kurzanzeige]

Set 1 | Setgröße 234

1	*Tsv	j Zeitschlagwort
2	*Tsv	j Slogan
3	*Tsv	j Soggettario per i cataloghi delle biblioteche italiane
4	*Tsv	j Geographisches Schlagwort
5	*Tsv	j Formschlagwort
6	*Tsv	j Descriptor
7	*Tsv	j Schlagwort
8	*Tsv	j Schlagwortkatalog
9	*Tsv	j Schlagwort <Parole>
10	*Tsv	j Politisches Schlagwort
11	*Tsv	j Schlagwortregister
12	*Tsv	j Stichwort
13	*Tsv	j Schlagwortnormdatei

Anwendungsgebiet: Verschlagwortung

Kritikpunkte I

- Mehrere Ansetzungsformen, aber trotzdem ist exakte/trunkierte Suche nicht immer ausreichend

Beispiele:

- HD-DVD
 - ???- Stokes-Gleichung? / -equation?
- **Erweiterung der Suchmöglichkeiten ist sinnvoll**



Anwendungsgebiet: Verschlagwortung

Kritikpunkte II

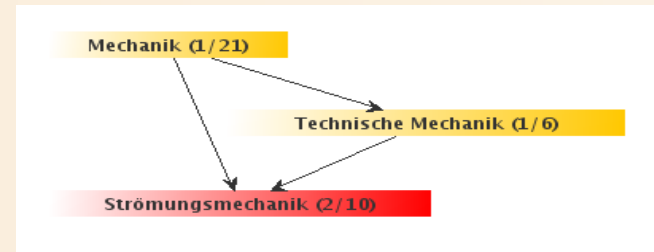
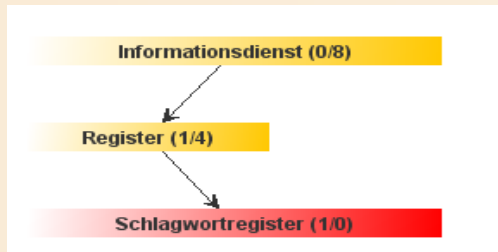
- Suche nach Unterbegriffen ist aufwändig / fehleranfällig (nur in Suchmenge)
- Sortierung der Schlagworte nach Ansetzungsdatum, nicht nach Beziehungen



➤ Hierarchische Darstellung und Navigation

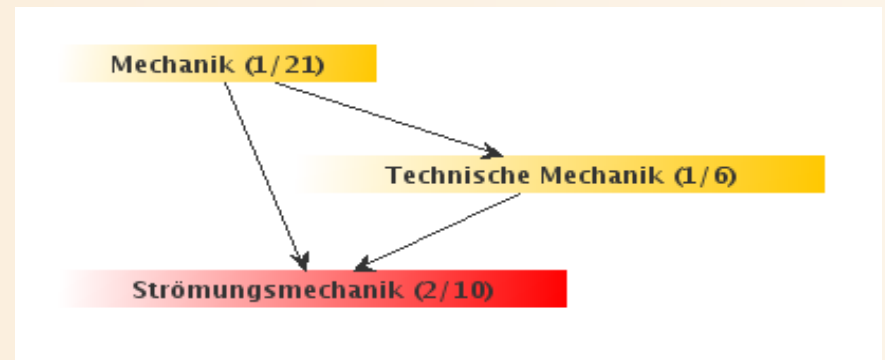
Anwendungsgebiet: Ansetzung

- Ansetzung neuer Schlagworte oft in Analogie zu bestehenden gewünscht
- **Gleichzeitige Darstellung** mehrerer Schlagworte interessant
- **Copy & Paste-Funktionalität** wünschenswert



Anwendungsgebiet: Ansetzung

- Datensätze: Nur eine Oberbegriffsebene wird spezifiziert
- Charakteristischer Fehler: Oberbegriff/Schlagwort haben denselben Oberbegriff
- **Hierarchische Visualisierung** kann diese Fehler „auf einen Blick“ zeigen



Nebeneffekt: Datenpflege

- Automatische Analyse der SWD zeigte verschiedene Fehler
- Mittlerweile Korrektur von
 - mehreren hundert Dubletten
 - 100+ unsauber definierten Einträgen
 - Verschiedenen Fehlern in der Vererbungshierarchie („Zirkel“)

SWD-Explorer

- Software zur
 - **erweiterten Suche** und
 - **grafischen Navigation**in der Schlagwortnormdatei (SWD)

- Entstanden im Rahmen einer Masterarbeit am Institut für Bibliothekswissenschaften Berlin
[urn:nbn:de:kobv:11-100112972](https://nbn-resolving.org/urn:nbn:de:kobv:11-100112972)

Features

- Schnelles Verständnis durch direktes Wahrnehmen struktureller/hierarchischer Beziehungen
- Effiziente Suche spezifischer Unterbegriffe durch mausgesteuerte Navigation
- Vermeidung hierarchiebasierter Fehler

Features

- Beschleunigtes Ansetzen neuer Schlagworte durch vereinfachte Einbeziehung vergleichbarer Schlagworte („copy - paste“ / parallele Darstellung)
- Komfortable Suche auf allen Ansetzungsformen:
 - exakt
 - trunziert (rechts-)
 - erweiterte Bool'sche Suche (und/oder)
 - Reguläre Ausdrücke (Inline-Trunkierung!)
 - Suche nach Datensatz (GVK-PPN)

Features

- Implementierung einer Suchhistorie
- Originaldatensatz (Pica3) jederzeit anzeigbar, daher hohe Informationsdichte
- Jedes Schlagwort wird je Suchanfrage nur einmal angezeigt

Technik

- Eigenentwicklung / Open Source, daher kostenlose Weitergabe möglich
- Modularer Aufbau für einfache Erweiterbarkeit
- Plattformunabhängige Programmierung in Java 6



Technik

- „Down to earth“-Ansatz: Sugiyama-Algorithmus (1981), Hash-Indizierung
- Offline per Datenbankabzug
- Verbesserungsspielraum... ;-)



Programmdemonstration

Visualisierung der GND...?

Wichtigste Unterschiede

- Zusammenlegung (= mehr Typen von Entitäten, größerer Umfang)
- Geändertes Datenformat
- Wegfall mehrgliedriger Oberbegriffe



Visualisierung der GND...?

Abgrenzung der Datenmenge

- **Frage:** Welche Datenmenge als neue Grundlage?
- Bisher: ca. 160.000 Normdatensätze



Visualisierung der GND...?

Abgrenzung der Datenmenge

- **Frage:** Welche Datenmenge als neue Grundlage?
- **Alternative 1: Konservativer Ansatz**
 - Eingrenzung auf Sachschlagworte
 - Mittels Entitätencodes (saz, sie, sih, sis,...) :
z.Zt. 54.000 Datensätze
 - Mittels Satzart/Eingrenzung auf Ts* :
z.Zt. 325.000+ Datensätze



Visualisierung der GND...?

Abgrenzung der Datenmenge

- **Frage:** Welche Datenmenge als neue Grundlage?
- **Alternative 2: Vollständiger Ansatz**
 - Alle Normdaten („GND-Explorer“)
 - Vorteil: Darstellung komplexer Relationen
 - Aber: Hierarchische Darstellung nur in Bezug auf Sachschlagworte sinnvoll
 - Extreme Datenmenge

Visualisierung der GND...?

Umgang mit Relationen?

- 132 Relationen, davon viele neu
- Viele Beziehungen zu Personen, Körperschaften etc.
- Konservativer Weg (auch bisher ungelöst):
vbal („Verwandeter Begriff allgemein“) sollte mindestens dargestellt werden, da analog zu „Schwesterknoten“

Visualisierung der GND...?

GND: Liste der Codierungen in Unterfeld \$4¹ – alphabetisch nach Code

Stand: 01. Dezember 2011, überarbeitet: 03. Mai 2012²

Nr.	Code ^{3 4}	Relation	MARCTerm
1.	abku	Abkuezung	Abkuezung
2.	adel	Adelstitel	Adelstitel
3.	adre	Adressat	Adressat
4.	adue	administrative Ueberordnung	Ueberordnung
5.	affi	Affiliation	Affiliation
6.	akad	Akademischer Grad	Akademischer Grad
7.	akti	Taetigkeitsbereich	Taetigkeitsbereich
8.	anla	Anlass	Anlass
9.	anno	Annotator	Annotator
10.	arch	Architekt	Architekt
11.	arra	Arrangeur	Arrangeur
12.	aust	Aussteller	Aussteller
13.	aut1	Verfasserschaft, erste	Verfasserschaft1
14.	auta	Verfasserschaft	Verfasserschaft
15.	autf	Fiktiver Verfasser	Fiktiver Verfasser
16.	autn	Verfasser, zugeschrieben	Zugeschriebener Verfasser

Weiterentwicklung

- „GND-Pause“ vorbei
- Umstieg auf Marc & GND-Nummer (statt Pica3 und GBV-PPN)
- Bessere Methode zur Datengewinnung (unAPI? DNB-OAI? Online-Tool?)



Weiterentwicklung

- Optimierung des Interfaces
- Steigerung der Effizienz durch DBS oder Suchmaschinentechnologie
 - Unschärfe Suche



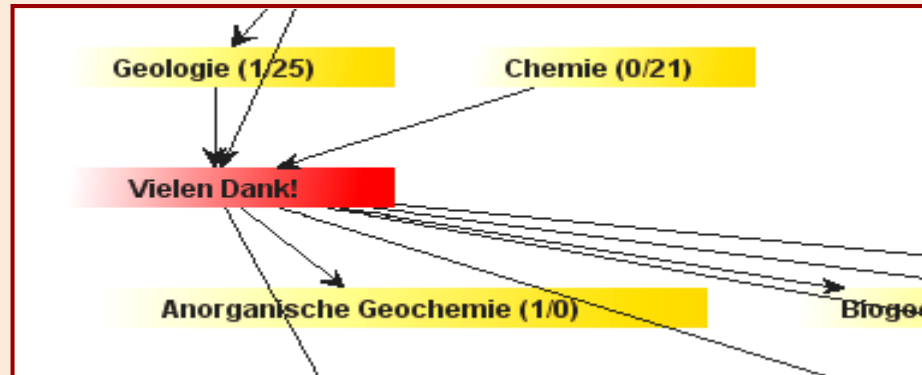
Weiterentwicklung

- Integration in Weboberflächen
 - Hierarchische Navigation in Weboberflächen
 - Ergänzung/Erweiterung facettierter Suche?

The screenshot shows a search interface with the following elements:

- Ihre Suche**: A search bar containing the query 'semantic web' with a red 'x' icon to clear it. An RSS icon and the text 'RSS Was ist das?' are in the top right corner.
- Suche eingrenzen**: A section for refining the search with the following options:
 - Verfügbarkeit:
 - Alle Bibliotheken:
 - Eingrenzen** (button)
- Faceted Search**: A list of filters with expandable/collapsible arrows:
 - Verfasser
 - Format
 - Sprache
 - Erscheinungsjahr (expanded):
 - [2008-2009](#) (42)
 - [2006-2007](#) (69)
 - [2004-2005](#) (59)
 - [2002-2003](#) (30)
 - [2000-2001](#) (1)
 - Thema (grob) (expanded):
 - [Informatik](#) (169)
 - [Bibliotheks- und Informationswissenschaft](#) (66)
 - [Industrielle Fertigung](#) (52)
 - [Ingenieurwissenschaften](#) (52)
 - [Mathematik](#) (22)
- [mehr...](#)

Ideen?



maas@sub.uni-hamburg.de