

Fremddatenanalyse für computerunterstützte Inhaltserschliessung

9. Mai 2017, Lukas Fischer



significance

excellence

*Masterarbeit von Ursula Jud-Reichlen
MAS in Bibliotheks- und Informationswissenschaften
Universität Zürich, 2017*

*Referent: Dr. Peter Schäuble
Eurospider Information Technology AG*

*Auswertungen durchgeführt von Lukas Fischer
Praktikant bei Eurospider Information Technology AG*

significance

excellence

Inhaltsverzeichnis

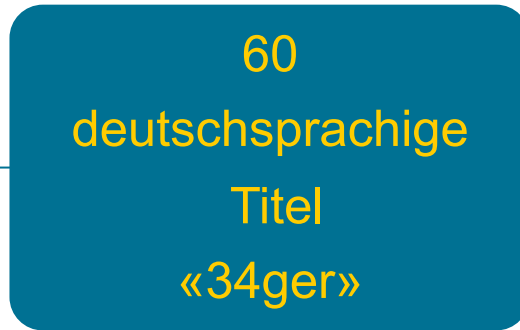
- *Übersicht*
- *Verwendete Daten und Methoden*
- *Evaluation Fremddatenimport*
- *Evaluation approximative Übersetzungen*
- *Evaluation Ähnlichkeitssuche*
- *Fazit*

technology



Übersicht (Entwurf)

Computerunterstützte Inhaltserschliessung		Zielvokabular	Merkmale (Features)	Methoden	Evaluation
DA-2	FDI	GND	ISBN	Vergleich ISBN	[Jud, 2017]
	Approximative Übersetzung		Fremdvokabulare	div. Konkordanz- tabellen	
	Ähnlichkeits- Suche		Textuelle Features	Regressions- methode	
DNB	Klassifikato- rische IE	DNB- Sachgruppe	Textuelle Merkmale	SVM	[Mödden/ Tomanek, 2012]
	Verbale IE	GND	Thesaurus	regelbasiert	[Uhlmann, 2013]
FRED		GND	ISBN	Vergleich ISBN	
UB Bern		ISDC- Klassifikation	UB Bern Hausklassifikation	Konkordanztabelle HK-ISDC	
ZBW		GND	Konkordanztabelle STW-GND	Probabilistische Methode/PLSA	[Groß, 2010]





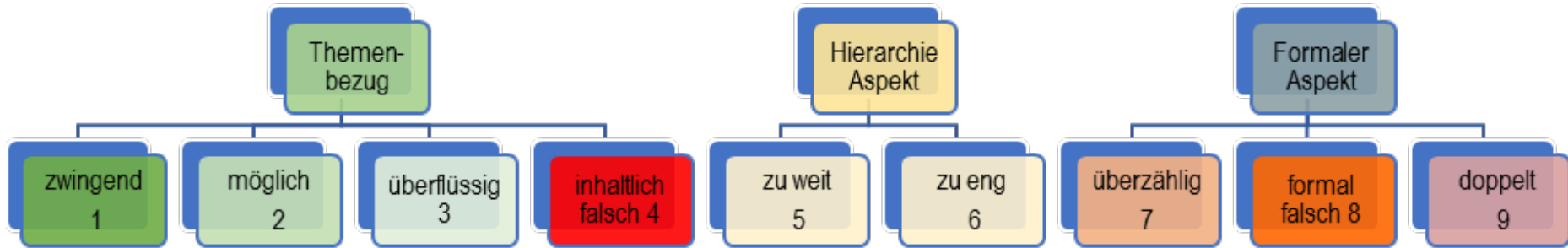
Datenbereitstellung für Goldstandard

ID:	414662822
Title:	Schadensersatz und Entschädigung in Folge von Diskriminierung im deutschen und englischen Arbeitsrecht : eine rechtsvergleichende Studie unter Berücksichtigung des Europarechts / Mathias Kaufmann
Author:	Kaufmann, Mathias
Language:	ger
Content:	http://swbplus.bsz-bw.de/bsz414662822inh.htm
Keyword:	650 7 \$a Arbeitsverhältnis \$0 (DE-588)4002799-5 \$2 gnd
Keyword:	650 7 \$a Diskriminierung \$0 (DE-588)4012472-1 \$2 gnd
Keyword:	650 7 \$a Rechtsvergleich \$0 (DE-588)4115712-6 \$2 gnd
Keyword:	650 7 \$a Schadensersatzanspruch \$0 (DE-588)4116364-3 \$2 gnd
Keyword:	651 7 \$a Deutschland \$0 (DE-588)4011882-4 \$2 gnd
Keyword:	651 7 \$a England \$0 (DE-588)4014770-8 \$2 gnd
ID:	400125412
Title:	Immobilien in den USA : Ratgeber für alle zukünftigen Wohneigentümer / Kai Blum
Author:	Blum, Kai (1969-)
Language:	ger
Content:	http://swbplus.bsz-bw.de/bsz400125412inh.htm
Keyword:	650 7 \$a Grundstückskauf \$0 (DE-588)4022360-7 \$2 gnd
Keyword:	650 7 \$a Wohneigentum \$0 (DE-588)4117670-4 \$2 gnd
Keyword:	650 7 \$a Kauf \$0 (DE-588)4128692-3 \$2 gnd
Keyword:	651 7 \$a USA \$0 (DE-588)4078704-7 \$2 gnd
Keyword:	655 7 \$a Ratgeber \$2 gnd

- ID/PPN
- Titel
- Autor
- Sprache
- Link zum Inhaltsverzeichnis
- Schlagwörter aus Eigen- und Fremddaten (Herkunft nicht ersichtlich)



Drei-Dimensionen-Schema nach Jud



1 zwingend «muss»

2 möglich «kann», evtl. spezifischerer Begriff

3 überflüssig; aber nicht falsch, jedoch sehr allgemein oder nur auf einen Teilaspekt «subgoal» bezogen

4 inhaltlich falsch

5 zu weit (hierarchischer) Unterbegriff würde es geben und wäre vorzuziehen

6 zu eng (hierarchischer) Oberbegriff würde es geben und wäre vorzuziehen

7 überzählig aufgrund eines Regelverstosses (Unter- u. Oberbegriff zusammen verwendet oder Verstoß gegen Anweisungen eines Hinweissatzes)

8 formal fehlerhaft (Zusatz «content» fehlt, Kongress anstelle Konferenzschrift, Konferenzschrift ohne Ort/Jahr)

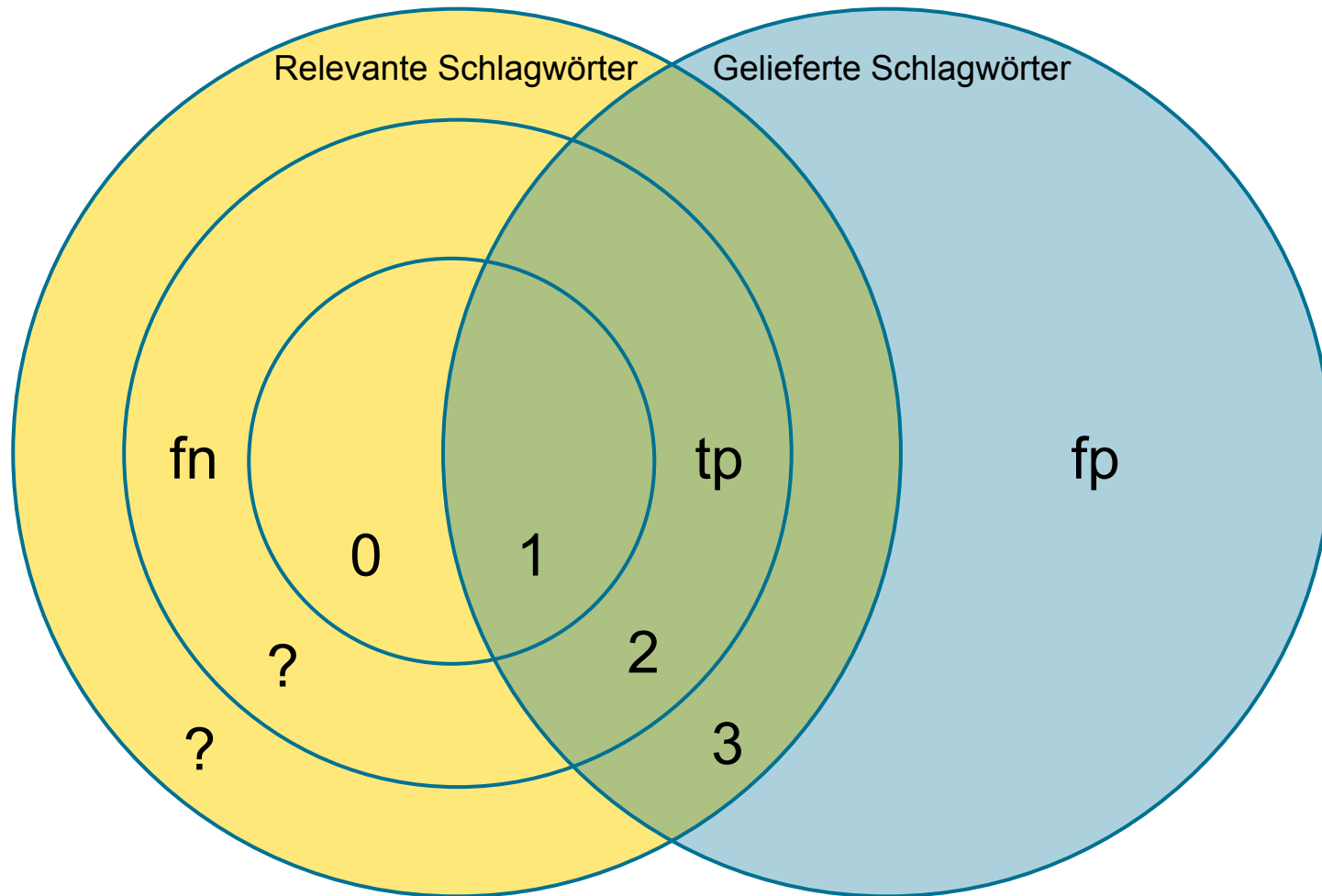
9 doppelt (überzählig aufgrund des Formatwechsels und formal fehlerhaft, siehe [8])



Ermittlung des Goldstandards

ID:	414662822	
Title:	Schadensersatz und Entschädigung in Folge von Diskriminierung im deutschen und englischen Arbeitsrecht : eine rechtsvergleichende Studie unter Berücksichtigung des Europarechts / Mathias Kaufmann	
Author:	Kaufmann, Mathias	
Language:	ger	
Content:	http://swbplus.bsz-bw.de/bsz414662822inh.htm	
Keyword:	650 7 \$a Schadensersatzanspruch \$0 (DE-588)4116364-3 \$2 gnd	3
Keyword:	651 7 \$a Deutschland \$0 (DE-588)4011882-4 \$2 gnd	4
Keyword:	650 7 \$a Arbeitsverhältnis \$0 (DE-588)4002799-5 \$2 gnd	3
Keyword:	650 7 \$a Diskriminierung \$0 (DE-588)4012472-1 \$2 gnd	1
Keyword:	650 7 \$a Rechtsvergleich \$0 (DE-588)4115712-6 \$2 gnd	1
Keyword:	651 7 \$a England \$0 (DE-588)4014770-8 \$2 gnd	1
Keyword:	610 1 \$a Deutschland \$t Allgemeines Gleichbehandlungsgesetz \$0 (DE-588)7542750-3 \$2 gnd	0
Keyword:	650 \$a Arbeitsrecht \$0 (DE-588)4002769-7 \$2 gnd	0
Keyword:	655 7 \$a Hochschulschrift \$0 (DE-588)4113937-9 \$2 gnd-content	0

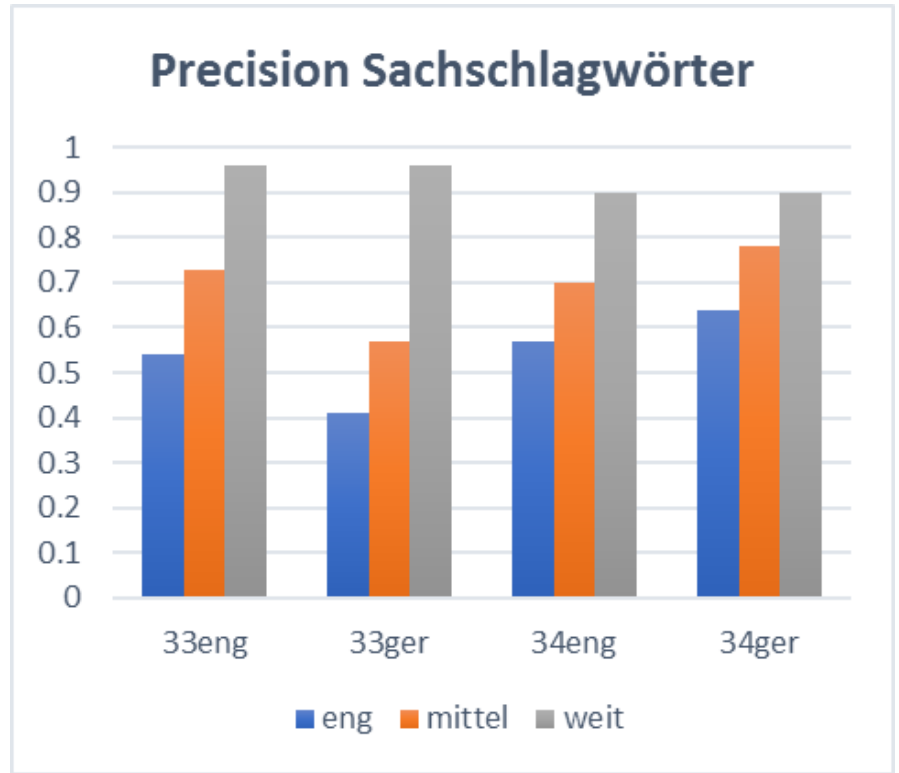
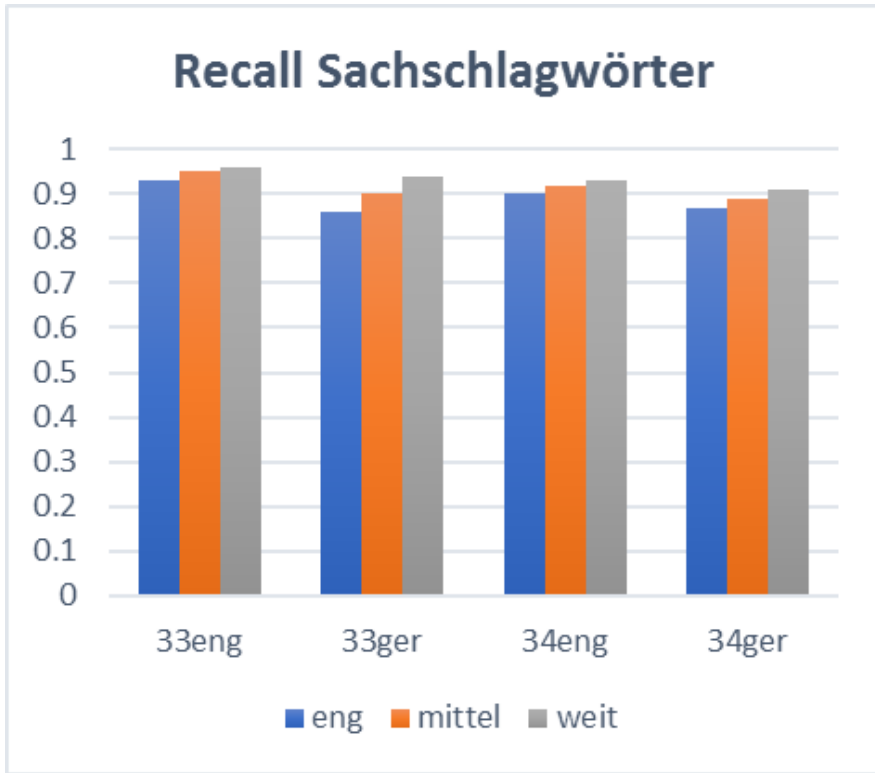
- Annotation gemäss «Drei-Dimensionen-Schema»
- Fehlende zwingende Schlagwörter werden manuell hinzugefügt und mit «0» annotiert.
- Fehlende mögliche [2] und fehlende überflüssige, aber nicht falsche [3] Schlagwörter werden nicht hinzugefügt.



$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

- Annotationen 1, 2 und 3 sehr oft genutzt (sachlich nicht falsch).
- Annotation 0 oft genutzt (fehlend):
 - häufig bei deutschen Titeln.
 - überdurchschnittlich oft bei Formschlagwörtern.
- Annotation 4 sehr selten genutzt (sachlich falsch):
 - Schlussfolgerung: Schlagwörter sind grundsätzlich von guter Qualität.
- Annotationen 5, 6 und 7 selten genutzt (formal fehlerhaft).
- Zahlreiche Annotationen 8 und 9, oft auch in Kombination (dito).



- Automatische Übernahme von gewissen Daten aus der Formalerschliessung.
- Automatische Generierung von gewissen Formalschlagwörtern durch Erkennung textueller Merkmale in vorhandenen Kataloganreicherungen bzw. in **digitalisierten Vorwörtern**:
 - Hochschulschriften:
Vorwort: «Die vorliegende **Dissertation** ist an zwei **Universitäten** entstanden».
 - Lehrbuch:
Inhaltsverzeichnis: «Kapitel 6: Lösungen zu den **Übungsaufgaben**».
 - Aufsatzsammlung:
Vorhandensein von mehreren Personennamen und verschiedenen Sprachen im Inhaltsverzeichnis.
 - Konferenzschriften:
Vorwort: «Der Vorliegende Band geht auf eine **Tagung** zurück, die im **Februar 2015** an der Universität **Hamburg** abgehalten wurde.»

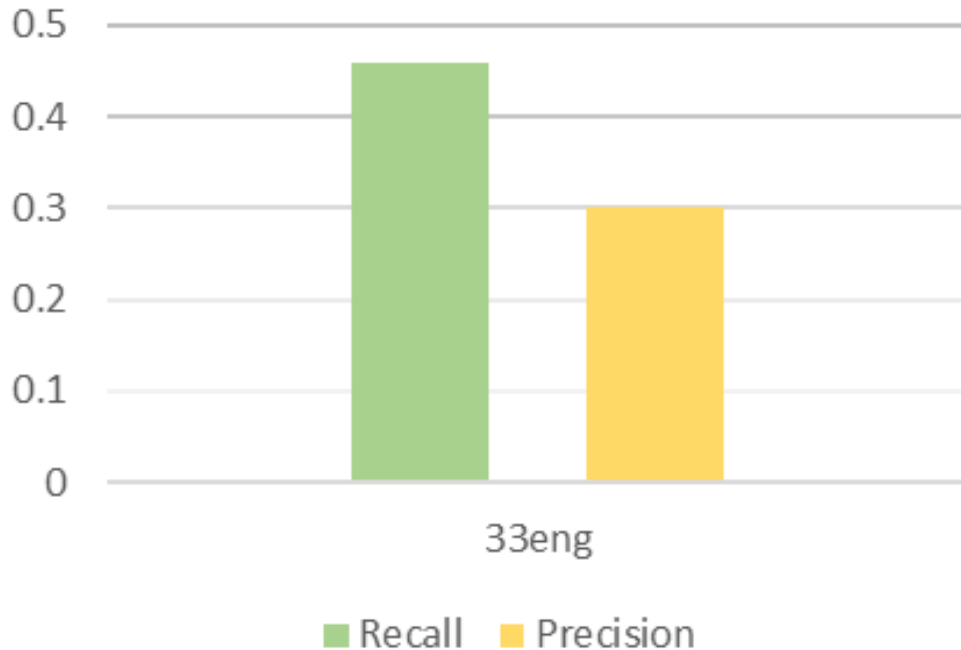
- Fremddatenquellen: BVB, GVK, IDSBB, OBV und WorldCat
- Schlagwörter aus fremden Erschliessungssystemen: LCSH, MeSH, STW, GBV, RERO, Rameau und GTT
- Übersetzungen mittels Konkordanzen:
 - Nicht-GND-Schlagwörter werden in GND-Schlagwörter übersetzt.



Vorschläge				
+ Amt	GND	Sach	GVK:LCSH, WC:LCSH, WC:RAM	
+ Behörde	GND	Sach	GVK:LCSH, WC:LCSH, WC:RAM	
+ Börsenkrach	GND	Sach	GVK:LCSH, WC:LCSH	
+ Dienststelle	GND	Sach	WC:RAM	
+ Elektronisches Buch	GND	Sach	WC:LCSH	
+ Finanzkrise	GND	Sach	BVB, IDSBB, OBV, WC	
+ Finanzpolitik	GND	Sach	WC	
+ Gemeindeverwaltung	GND	Sach	GVK:LCSH, WC:LCSH, WC:RAM	
+ Haushaltsdefizit	GND	Sach	BVB, IDSBB, OBV, WC	
+ Hoheitsverwaltung	GND	Sach	GVK:LCSH, WC:LCSH, WC:RAM	
+ Innovation	GND	Sach	WC	
+ Kommunalpolitik	GND	Sach	WC	
+ Konsolidierung	GND	Sach	BVB, IDSBB, OBV, WC	
+ Kurssturz	GND	Sach	GVK:LCSH, WC:LCSH	
+ Lokale Machtstruktur	GND	Sach	GVK:LCSH, WC:LCSH	
+ Regionalpolitik	GND	Sach	BVB, IDSBB, OBV, WC	
+ Schlüsselzuweisung	GND	Sach	GVK:LCSH, WC:LCSH, WC:RAM	
+ Städtebund	GND	Sach	GVK:LCSH, WC:LCSH	
+ Verwaltung	GND	Sach	GVK:LCSH, WC:LCSH, WC:RAM	
+ Öffentliches Amt	GND	Sach	GVK:LCSH, WC:LCSH, WC:RAM	
+ Europa	GND	Geo.	BVB, GVK:LCSH, IDSBB, OBV, WC, WC:LCSH	
+ Aufsatzsammlung	GND	Form	BVB, IDSBB, OBV	
Electronic books	LCSH:PART	Sach	WC	
Financial crises	LCSH:PART	Sach	GVK, WC	
Local finance	LCSH:PART	Sach	GVK, WC	
Local government	LCSH:PART	Sach	GVK, WC	
Public administration	LCSH:PART	Sach	GVK, WC	
Europe	LCSH:PART	Geo.	GVK, WC	
Electronic books	LCSH	Sach	WC	
Financial crises—Europe	LCSH	Sach	GVK, WC	
Local finance—Europe	LCSH	Sach	GVK, WC	



APPROXIMATIVE ÜBERSETZUNGEN



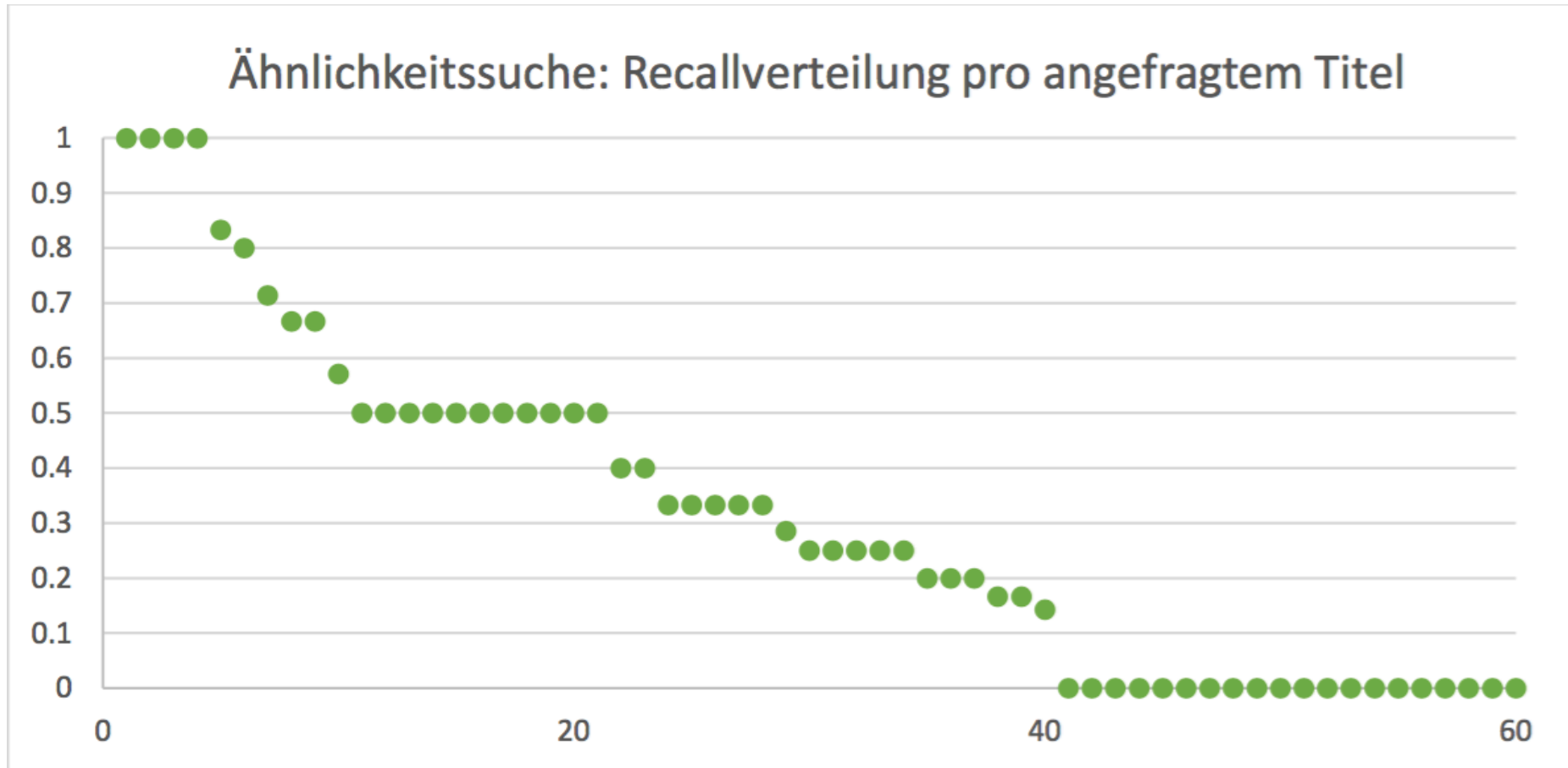
- Auswertungsergebnisse entsprechen nicht dem subjektiven Empfinden der Untersucherin: Viele gute Schlagwortvorschläge!
- Kann der persönlichen Fortbildung dienen.
- Untersuchungsmethode retrospektiv eher als ungünstig beurteilt.



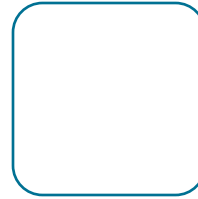
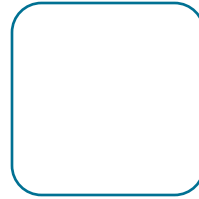
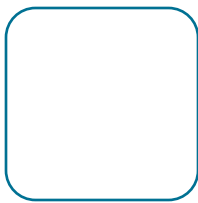
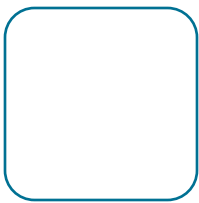
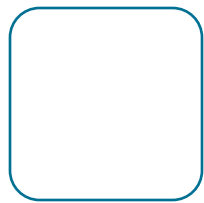
Evaluation approximative Übersetzungen Kategorie «33eng»

Vorschläge			
+ Amt	GND	Sach	GVK:LCSH, WC:LCSH, WC:RAM
+ Behörde	GND	Sach	GVK:LCSH, WC:LCSH, WC:RAM
+ Börsenkrach	GND	Sach	GVK:LCSH, WC:LCSH
+ Dienststelle	GND	Sach	WC:RAM
+ Elektronisches Buch	GND	Sach	WC:LCSH
+ Finanzkrise	GND	Sach	BVB, IDSBB, OBV, WC
+ Finanzpolitik	GND	Sach	WC
+ Gemeindeverwaltung	GND	Sach	GVK:LCSH, WC:LCSH, WC:RAM
+ Haushaltsdefizit	GND	Sach	BVB, IDSBB, OBV, WC
+ Hoheitsverwaltung	GND	Sach	GVK:LCSH, WC:LCSH, WC:RAM
+ Innovation	GND	Sach	WC
+ Kommunalpolitik	GND	Sach	WC
+ Konsolidierung	GND	Sach	BVB, IDSBB, OBV, WC
+ Kurssturz	GND	Sach	GVK:LCSH, WC:LCSH
+ Lokale Machtstruktur	GND	Sach	GVK:LCSH, WC:LCSH
+ Regionalpolitik	GND	Sach	BVB, IDSBB, OBV, WC
+ Schlüsselzuweisung	GND	Sach	GVK:LCSH, WC:LCSH, WC:RAM
+ Städtebund	GND	Sach	GVK:LCSH, WC:LCSH
+ Verwaltung	GND	Sach	GVK:LCSH, WC:LCSH, WC:RAM
+ Öffentliches Amt	GND	Sach	GVK:LCSH, WC:LCSH, WC:RAM
+ Europa	GND	Geo.	BVB, GVK:LCSH, IDSBB, OBV, WC, WC:LCSH
+ Aufsatzsammlung	GND	Form	BVB, IDSBB, OBV
Electronic books	LCSH:PART	Sach	WC
Financial crises	LCSH:PART	Sach	GVK, WC
Local finance	LCSH:PART	Sach	GVK, WC
Local government	LCSH:PART	Sach	GVK, WC
Public administration	LCSH:PART	Sach	GVK, WC
Europe	LCSH:PART	Geo.	GVK, WC
Electronic books	LCSH	Sach	WC
Financial crises—Europe	LCSH	Sach	GVK, WC
Local finance—Europe	LCSH	Sach	GVK, WC

Die Ähnlichkeitssuche des DA-2 zeigt für einen angefragten Titelsatz bis zu 40 ähnliche Titel an. Aus der Beschlagwortung dieser Titel können Vorschläge für die Beschlagwortung des angefragten Titels gewonnen werden.



- DA-2:
 - Fremddatenimport funktioniert gut.
 - Qualität der Fremddaten ist gut.
 - Funktion approximative Übersetzungen erzeugt zusätzliche gute Schlagwortvorschläge.
 - DA-2 eignet sich für eine kooperative Erschliessung.
- Generell:
 - Weiternutzungsmöglichkeiten von Erschliessungsdaten beachten; Kooperationen und Absprachen sind wichtig und wertvoll.
 - Kataloganreicherungen sind wertvoll, je mehr, desto besser; man weiss nie, was später nützlich sein könnte...



*Besten Dank für Ihre
Aufmerksamkeit*

Für Fragen stehe ich ihnen jetzt gerne zur Verfügung.

